AD-A118 444 TEXAS INSTRUMENTS INC DALLAS F/G 17/2
VOICE VERIFICATION UPGRADE.(U)
JUN 82 R L DAVIS, J T SINNAMON, D L COX F30602-79-C-0190
UNCLASSIFIED TI-08-82-07 RADC-TR-82-139 NL

1 of 3
AD
A118444

RADC-TR-82-159
Final Technical Report
June 1982

VOICE VERIFICATION UPGRADE

Texas Instruments Incorporated

Robert L. Davis, James T. Sinnamon and David L. Cox

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-82-139 has been reviewed and is approved for publication.

APPROVED: *Richard S. Vonusa*

RICHARD S. VONUSA
Project Engineer

APPROVED: *A. J. Driscoll*

A.J. DRISCOLL, Colonel, USAF
Chief, Intelligence and Reconnaissance Division

FOR THE COMMANDER: *John P. Huss*

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA), Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-82-139	2. GOVT ACCESSION NO. AD-A118 444	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) VOICE VERIFICATION UPGRADE		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 12 May 80 - 31 Dec 81
		6. PERFORMING ORG. REPORT NUMBER 08-82-07
7. AUTHOR(s) Robert L. Davis James T. Sinnamon David L. Cox		8. CONTRACT OR GRANT NUMBER(s) F30602-79-C-0190
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas Instruments Incorporated 13500 North Central Expressway Dallas TX 75265		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63714F 681E0903
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss AFB NY 13441		12. REPORT DATE June 1982
		13. NUMBER OF PAGES 254
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Same		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Richard S. Vonusa (IRAA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speaker Verification
Voice Authentication
Speaker Recognition
Speech Recognition
Entry Control

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This contract had two major objectives. The first was to build, test, and deliver to the government an entry control system using speaker verification (voice authentication) as the mechanism for verifying the user's claimed identity. This system included a physical mantrap, with an integral weight scale to prevent more than one user from gaining access with one verification (tailgating). The speaker verification part of the entry control system contained all the updates and embellishments to the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

algorithm that was developed earlier for the BISS (Base and Installation Security System) system under contract with the Electronic Systems Division of the USAF. These updates were tested prior to and during the contract on an operational system used at Texas Instruments in Dallas, Texas, for controlling entry to the Corporate Information Center (CIC). Rather than update the existing BISS-ASV-ADM (BISS - Automatic Speaker Verification - Advanced Development Model), the complete, updated algorithm was provided to the Air Force on a totally new system of three computers, which was tested for six months at Texas Instruments. Over 13,000 accesses were performed using this system, with less than 1.0% of the users being refused access based on their vocal characteristics (0.75% if users limited to two attempts). Off-line tests of casual impostors yielded an error rate of less than 1.0% with an over 90% confidence level. This Voice Verification Upgrade (VVU) system has been delivered and installed at Rome Air Development Center, Griffiss Air Force Base, New York, and is operational in the RADC/IRAA Laboratory.

The second purpose of this contract was the continued research into voice authentication algorithms and entry control system performance. Pursuant to these objectives, the following studies were performed:

1. a trade-off study on speaker verification performance as a function of the prompting words,

2. a simulation of booth traffic for an entry control system using speaker verification, and

3. a study of speaker verification performance using an LPC-based prediction residual.

The last of these three studies was by far the most extensive, and provided an order of magnitude improvement in performance, resulting in performance exceeding that set as goals for this contract ($<$ 1.0% true speaker rejections and $<$ 0.1% impostor acceptances). Included in this study was the completion of an on-line, real-time demonstration of the LPC-based speaker verification method on the VAX 11/780 at the Speech Systems Research Laboratory at the Texas Instruments facility in Dallas.

TABLE OF CONTENTS

SECTION		PAGE
I	INTRODUCTION	1
II	HARDWARE	7
	A. Central Computer System Hardware	7
	B. Voice Processor Parts List	10
III	SOFTWARE	22
	A. Host Processor Software	22
	1. The Host Data Base	23
	2. The Host Tasks	26
	3. Host Scenarios	29
	B. Voice Processor Software	30
	1. Task to Communicate with the Terminal Processor	31
	2. Terminal Task	34
	3. Voice Processing	35
	4. Host Communications Tasks	37
	C. Terminal Processor Software	37
	1. Operating System	38
	2. Speech Processing	38
	3. Communications	38
	4. Portal Control	40
	5. Local Terminal Interface	41
	D. Downloading the Voice Processor and the Terminal Processor	41
IV	BISS VOICE AUTHENTICATION ALGORITHM MODIFICATIONS	43
	A. Word Set	43
	B. Speech Preprocessing	48
	1. Filter Bank Definitions	48
	2. Regression	50
	3. Normalization	52

TABLF OF CONTENTS

(Continued)

SECTION		PAGE
	4. Quantization.	53
	5. Energy.	
	C. Verification Processing.	59
	1. Point-Pair Error Calculation.	60
	2. Decision Strategy	62
	3. Decision Functions.	64
	4. Decision Parameters	65
	D. Other Modifications.	66
V	ALGORITHM MODIFICATION TESTING.	68
	A. Off-Line 11-Speaker Testing.	68
	1. The Data Set.	68
	2. Filter Bank Definitions	68
	3. Performance Measures.	70
	4. Results	70
	B. On-Line Operational Testing for 980-Based CIC System .	73
	1. The Data.	73
	2. System Modification Log	74
	3. Results of On-Line Operational Testing.	74
VI	TEST OF THE VOICE VERIFICATION UPGRADE SYSTEM	79
	A. "True Speaker" Testing	79
	B. "Casual" Impostor Testing.	83
VII	RESIDUAL ENERGY BASED SPEAKER VERIFICATION.	88
	A. Similarity Measure	88
	B. Modifications to the Input Autocorrelation Function. .	88
	C. Time Registration.	89
	D. Derivation of the Residual Error Decision Function . .	91
	E. The Experimental Data Set.	95
	F. Fixed Pattern Time Normalization Experiments	95

TABLE OF CONTENTS
(Continued)

SECTION		PAGE
F.	Nonlinear Time Warping Experiments	103
G.	Automatic Enrollment	107
H.	Expanded Data Set Experiments	109
	REFERENCES	114

APPENDIXES

APPENDIX I	ENTRY CONTROL POINT SIMULATION MODEL	117
APPENDIX II	PROMPTING WORD SELECTION TRADE-OFF STUDY	159
APPENDIX III	ENROLLMENT AND VERIFICATION ALGORITHMS	181

LIST OF TABLES

TABLE		PAGE
1	Host Processor Parts List	7
2	Voice Processor Parts List	10
3	Entry Control Booth Parts List	14
4	Terminal Processor Parts List	18
5	ASV System Comparisons Overall	44
6	ASV System Comparisons Speech Input	45
7	ASV System Comparisons Enrollment	46
8	ASV System Comparisons Verification	47
9	Phrase Prompting Order During Enrollment	48
10	Filter Bank Definitions	49
11	Regression Vectors	51
12	Word Set #1 Used in Determining Quantization Thresholds	53
13	Word Set #2 Used in Determining Quantization Thresholds	54
14	High Amplitude Filters for BISS Filter Bank	55

LIST OF TABLES
(Continued)

TABLE		PAGE
15	High Amplitude Filters for TVP/VVU Filter Bank.	56
16	Quantization Thresholds DSG Filter Bank Normalizing by Standard Deviation.	58
17	Conditions for Determining Quantization Thresholds.	59
18	Point-Pair Error Parameters	61
19	Average Expected Scanning Errors.	62
20	Sets of Decision Function Thresholds Used	65
21	Sets of Max/Min Scanning Errors Allowed	66
22	Filter Bank Definitions	69
23	Speaker Expected Errors	71
24	Performance Comparison of Processing Methods.	72
25	Log of Changes to the 980-Based CIC System.	75
26	On-Line True Speaker and Off-Line Impostor Results for CIC Entry Control Booth Using BISS-Type Preprocessing	77
27	On-Line True Speaker and Off-Line Impostor Results for CIC Entry Control Booth Using IPMOD2-Type Preprocessing	78
28	Not Verifieds Across the User Population.	80
29	Performance Improvement for Eight Reenrolled Speakers	80
30	Performance as a Function of Number of Prior Verifications for the 990-Based Systems	81
31	Access Reject Rate (13,317 Accesses) as a Function of Number of Retries	82
32	VVU Performance by Sex.	83
33	Verification Performance Summary (Type 1 Analysis for VVU ASV System.	84
34	Casual Impostor Testing Results and Confidence Levels that True Error Rate is < 1.0% Based Upon Observed Error Rate. . .	85

LIST OF TABLES
(Continued)

TABLE		PAGE
35	Distribution of Impostor Successes.	87
36	Unregistered Phrase Rates for Impostors/True Speakers and Equal Error Rates in Percent.	101
37	Unregistered Phrase Rates for Impostors/True Speakers and Equal Error Rates in Percent.	103
38	Equal Error Rates (EERs) in Percent and Unregistered Phrase Rates for Impostors/True Speakers for a One-Phrase Decision Using a Nonlinearly Time Warped, Residual Energy Feature Vector.	110
39	Equal Error Rates in Percent for One-Phrase Decision Using Nonlinearly Time Warped, Residual Energy Feature Vector for Various Regression/LPC Order Combinations	111
40	Equal Error Rates in Percent as a Function of Number of Phrases for Nonlinearly Time Warped, Residual Energy Feature Vectors	112
41	True Speaker (TS) Rejects and Impostor (IM) Acceptances for a Full "CIC," Multiphrase Decision Strategy Using a Nonlinearly Time Warped, Residual Energy Decision Function	113

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1	Interaction of Government and Internal Funding for Voice Authentication at TI.	2
2	Texas Instruments' Digital Systems Group's Automatic Speaker Verification (ASV) System	4
3	Voice Verification Upgrade (VVU) System	6
4	VVU System Relationship	8

LIST OF ILLUSTRATIONS
(Continued)

FIGURE		PAGE
5	VVU System Components	9
6	Outside View of Entry Control Booth	12
7	View of User Terminal Mounted Inside Booth.	12
8	Layout of Equipment Bay in Entry Control Booth.	13
9	Terminal Processor and Peripheral Interface	15
10	Block Diagram of Speech I/O System	16
11	A Typical Analog Filter Board Filter Section.	17
12	Two Pole Active Bandpass Filter	19
13	Frequency Response Characteristics for Filter #1.	21
14	Sample User Performance Summary	24
15	Sample Portal Performance Summary	25
16	Inter-Task Communications for DSG's Voice Processor	32
17	Inter-Task Communications for the VVU System's Voice Processor	33
18	Quantization Threshold Comparisons.	57
19	CIC/VVU Verification Strategy Overview.	63
20	Comparison of BISS/DSG Preprocessing, Single-Phrase Performance for CIC	76
21	Observed Error Rate vs Sample Size to Insure A 90% Confidence Level that the True Error Rate < 1%.	86
22	Comparison of LPC Modelled Spectra With and Without Noise Floor	
23	Comparison of Basic and Modified Dynamic Programming Algorithms.	92
24	Speaker Verification Performance for Males Using Unnormalized Errors	97
25	Speaker Verification Performance for Females Using Unnormalized Errors	98

LIST OF ILLUSTRATIONS
(Continued)

FIGURE		PAGE
26	Speaker Verification Performance for Males Using Normalized Errors.	99
27	Speaker Verification Performance for Females Using Normalized Errors	100
28	Speaker Verification Performance Comparison Using Three Different LPC Residual Errors with Fixed Formats.	102
29	Speaker Verification Performance Comparison Using (1) Filter Bank, (2) LPC Residual Errors, and (3) LPC Residual Errors with Nonlinear Time Warping	104
30	Speaker Verification Performance Comparison Using (1) Filter Bank, (2) LPC Residual Errors, and (3) LPC Residual Errors with Nonlinear Time Warping	105

SECTION I

INTRODUCTION

This final report covers the sixth in a series of programs undertaken by Texas Instruments, under government sponsorship, to further develop speaker verification (voice authentication [1,2]) technology. The relationships between these six government-funded programs and internally funded voice authentication developments are shown in Figure 1. In the first program [3] (SV1), a promising high-performance speaker verification technology was developed and comprehensively tested in a laboratory environment, with accurate and reliable methods of time registration providing a major performance impact.

In the second program [4] (SV2), operationally important problems were solved to provide an operational capability for applications such as automatic entry control. Concurrent with this second program were:

> The development of an Advanced Development Model voice verification system for the Base and Installation Security Systems (BISS) program under Electronic Systems Division sponsorship [5] using a TI-980 minicomputer. (This system was subsequently tested by Mitre [6] in side-by-side tests of verification systems using handwriting and fingerprint.)
>
> The installation of an operational, fully automated entry control system, internally funded, to provide entry control to the Texas Instruments Corporate Information Center [7] (CIC), also using a TI-980 minicomputer.

In the third program [8] (SV3), advanced speech processing capabilities were developed to enhance speaker verification effectiveness and extension of speaker verification technology was made to other applications. Effort was focused on two specific applications: speaker verification using passwords embedded in free text and speaker identification (and subsequent verification) using spoken identification codes (called "Total Voice" verification). Both of these required major emphasis on the development of word recognition technology and the integration of recognition and verification techniques.

The fourth program [9] (Remote Terminal SV) was a study conducted to develop speaker verification techniques for use over degraded communication channels -- specifically telephone lines. A test of BISS type speaker verification technology was performed on a degraded channel and compensation techniques were then developed.

The fifth program [10] (Total Voice SV) was the coalescence of the Total Voice verification technology and the hardware of the Advanced Development Model BISS speaker verification system (then located at RADC) culminating in the installation of the Total Voice computer program on the BISS-SV system.

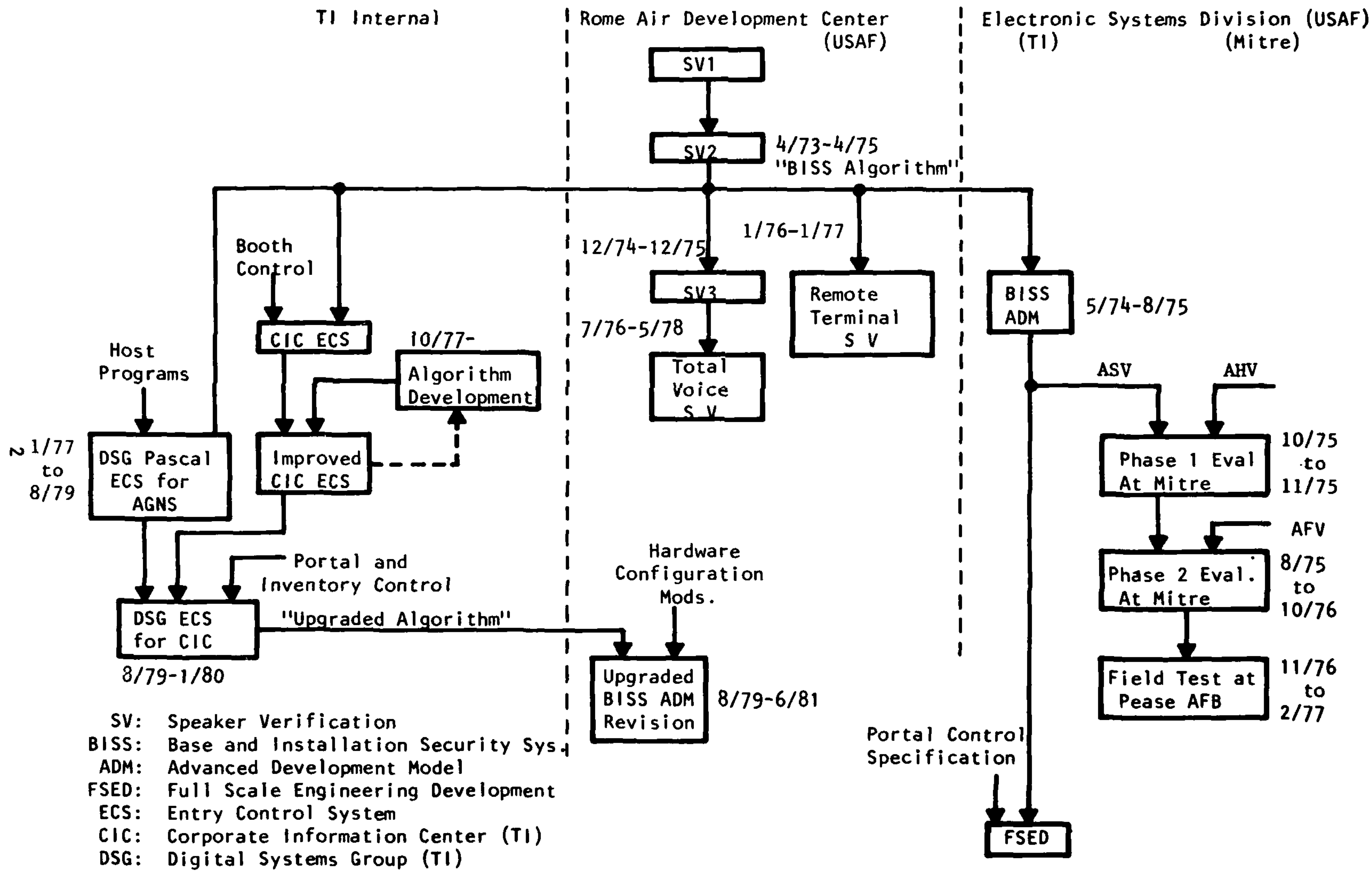


Figure 1 Interaction of Government and Internal Funding for Voice Authentication at TI

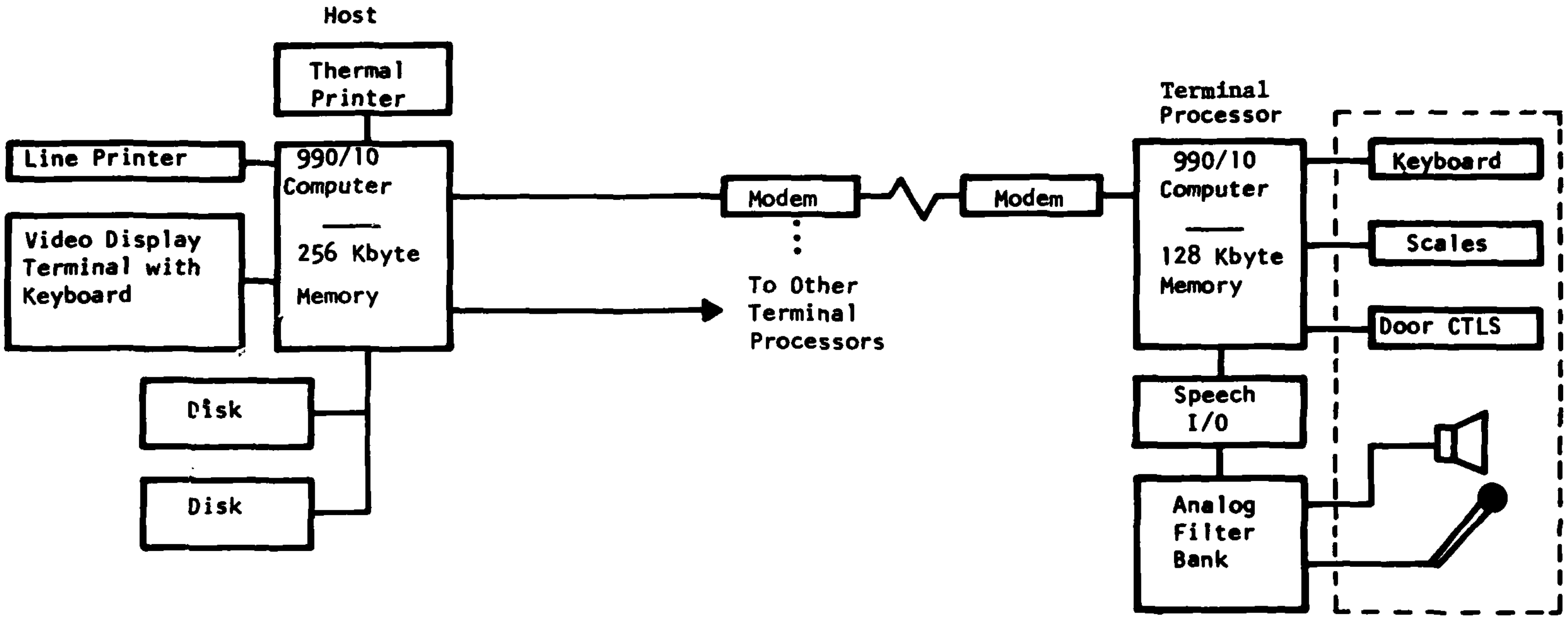
During the time period of the third, fourth and fifth contract, however, Texas Instruments was spending internal funds to improve the voice verification algorithms and to develop a commercially available entry control system using voice authentication. The algorithm development was done both on-line, using the entry control system (ECS) at CIC and off-line, using a specially collected laboratory data set.

In early 1977, TI's Digital Systems Group (DSG) began development of an entry control system using TI 990 minicomputers (rather than the TI 980's used in the older systems), with the programs being written primarily in Pascal, with a few 990 assembly language routines (rather than Fortran and 980 assembly language, as before). This entry control system used most of the algorithm improvements made during the trade-off studies on the CIC system, with the addition of more security protocol (inventory control, deactivation of users, time-of-day security checks, etc.), improved report generation capabilities, and an improved (more complicated) interface to the security personnel.

However, the need for such a system to be able to service multiple entrances, required changing the system hardware architecture, with some new limitations resulting from the introduction of the required digital communication line between the central computer and the entry control booth. (Analog signals from the speaker and to the microphone that were used when the computer was colocated with the booth are not appropriate for long distance transmission, as might be required for BISS, for example.) This digital communication line, although operated at 9600 bps, still provided a bottleneck, resulting in somewhat slower response time between the user's input speech and a verified/not verified response. In addition, the oral prompting used to direct the user could no longer be transferred over the communication line without further degrading the response time. (The prompts on the older systems were PCM-coded with 8,000, eight-bit samples per second, thus requiring about 8 seconds to transmit a one second prompt.) Hence, some efficient digital encoding scheme was required for the prompts, which would allow either transmission of the speech very efficiently, or efficient storage at the entry control booth end. Such coding, however, resulted in a degradation in the quality of the oral prompts.

The resulting system developed by DSG is shown in Figure 2. Such systems were delivered both to Allied General Nuclear Services (AGNS) in August 1979, and to CIC in April 1980, to replace the older, single booth, 980-based system.

Although this current contract (Voice Verification Upgrade [VVU]) effort between TI and RADC was initially intended to add the new algorithm improvements to the existing, 980-based, BISS-ASV-ADM, by the start of the contract effort, it was determined to be more to the government's benefit for Texas Instruments to deliver an entirely new ASV system, based on the newer, more serviceable, 990 minicomputers, taking advantage of the development effort by DSG. However, in order both to improve the response time of the system to the user (by sharing the computing load during a verification with a more powerful centrally located computer) and to reduce the cost of the computing equipment at the entry



Host Tasks

- o Security Protocol
- o Bookkeeping
- o Reference File Storage
- o Operator Interface
- o Inputting Personal Data During Enrollment
- o Impostor Data Storage
- o Report Generation
- o Operator Alert Via Thermal Printer

Terminal Processor Tasks

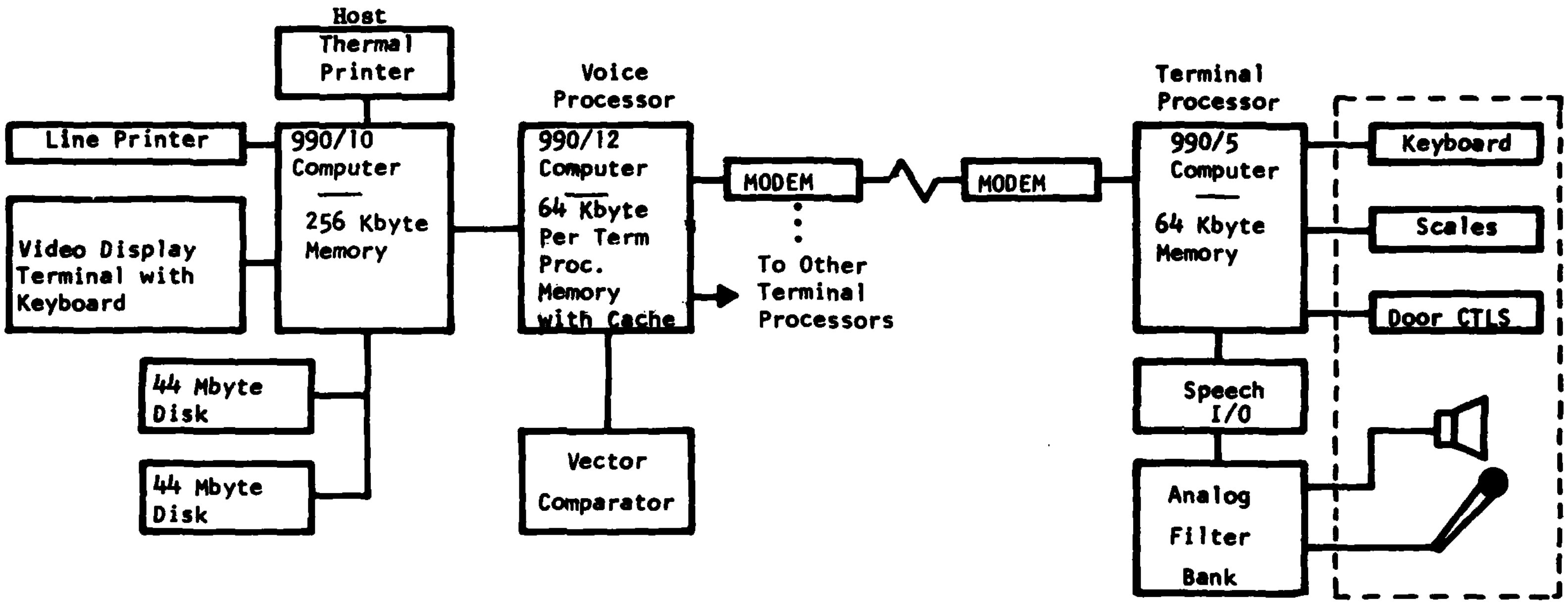
- o Input Speech Preprocessing and Compression
- o Phrase Registration
- o Verification
- o Reference File Updating
- o Synthesis of Auditory Prompting
- o Monitoring of Keyboard, Scales and Door Sensors
- o Control of Door Locks
- o Control of Tape Recorder

Figure 2 Texas Instruments' Digital Systems Group's Automatic Speaker Verification (ASV) System

control booth (eventually replacing it with a microcomputer), a modification was made to the system architecture being used by DSG, resulting in the system shown in Figure 3. This is the system delivered to RADC for this contract, and will be referred to as the VVU system throughout the remainder of this final report.

During late 1980, DSG developed and delivered a three portal system to AGNS that although similar to the system in Figure 2, did not contain any of the portal control (the DSG host communicated to an AGNS central security computer, which in turn controlled the booth doors and sensors), resulting in a computer at the entry control booth that only provided a verification decision to the host for a given reference file supplied by the DSG host. This architecture results in a stand-alone computer that makes a verification decision with an externally supplied reference file. This also is a viable method for achieving the desired cost reduction in the booth computer, and provides a method for supplying a product to customers already possessing a computer-controlled security system. However, due both to timing and to the BISS program's interest in a total security system, no modification to the system shown in Figure 3 was made.

In addition to the construction, evaluation and delivery of the VVU voice authentication system, a number of other experiments were performed: 1) a booth simulation using queing models, 2) experiments using LPC residual energy for speaker verification, and 3) a word trade-off study to determine verification performance as a function of spoken word. All these experiments are covered in this final report, in addition to a description of the VVU system and the results of its evaluation.



<u>Host Tasks</u>

- o Security Protocol
- o Bookkeeping
- o Reference File Storage
- o Operator Interface
- o Inputting Personal Data During Enrollment
- o Impostor Data Storage
- o Report Generation
- o Operator Alert via Thermal Printer

<u>Voice Processor Tasks</u>

- o Phrase Registration
- o Verification
- o Reference File Updating

<u>Terminal Processor Tasks</u>

- o Input Speech Preprocessing and Compression
- o Synthesis of Auditory Prompting
- o Monitoring of Keyboard, Scales and Door Sensors
- o Control of Door Locks
- o Control of Tape Recorder

Figure 3 Voice Verification Upgrade (VVU) System

SECTION II

HARDWARE

The upgraded BISS ADM ASV system is composed of two primary parts, a computer hardware system and an entry control booth. The relationship of the two is shown in Figure 4. A description of each major part of both the computer system and the entry control booth is given in the next two subsections.

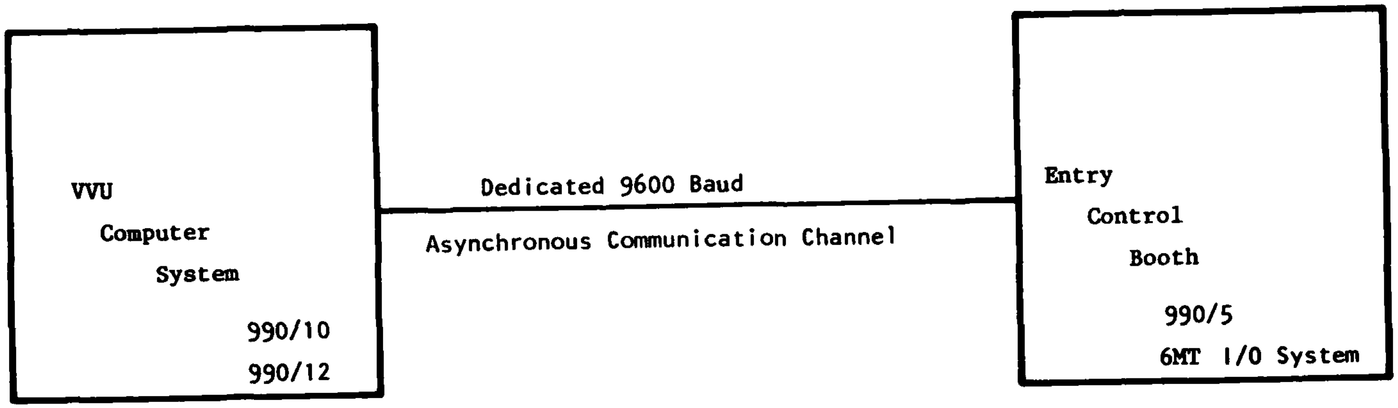
A. CENTRAL COMPUTER SYSTEM HARDWARE

The Computer System for the Upgraded BISS ADM ASV system is composed of a host processor, voice processor and terminal processor, as shown in Figure 5. The host and voice processors are colocated in a double bay desk with the terminal processor located in the equipment bay of the Entry Control Booth, which may be located remotely. The voice processor is connected via a 9600 baud asynchronous modem to the terminal processor. Terminal processor components will be discussed in the Entry Control Booth section.

Parts list for the colocated Host and Voice processors are given in Tables 1 and 2.

TABLE 1. HOST PROCESSOR PARTS LIST

	ITEM	QTY
1.	990/10 CPU, mem mapping, 13-slot chassis programmer panel, no standby power	1
2.	TILINE error-correcting memory subsystem (384K bytes)	1
3.	DS50 interface kit (for T50 disk drives) (includes 15 ft bus & 20 ft daisy chn cables)	1
4.	Trident daisy chain cable (15 ft)	1
5.	Trident radial cable (20 ft)	1
6.	T50 disk pack drive	2
7.	T50 disk drive terminator	1
8.	Low boy console for T50	2
9.	DS50 disk pack	4
10.	810 printer master kit (printer, interface, cable, paper tray and manuals)	1
11.	Model 810 printer stand, without paper tray	1
12.	743 KSR terminal	1
13.	743 KSR interface kit (includes 30 ft cable)	1
14.	911 VDT kit (dual 1920-character controller, 2 displays and keyboards)	1
15.	TILINE interface kit	1
16.	Rackmount equipment cabinet (also houses co-located voice processor)	1



- Operator interface and VVU system control point
- User and system disk file maintenance
- Verification/authentication decision making

- User interface - keyboard and speech prompting
- User voice data preprocessing
- Multi-user detection
- Door lock sensing and control

Figure 4 VVU System Relationship

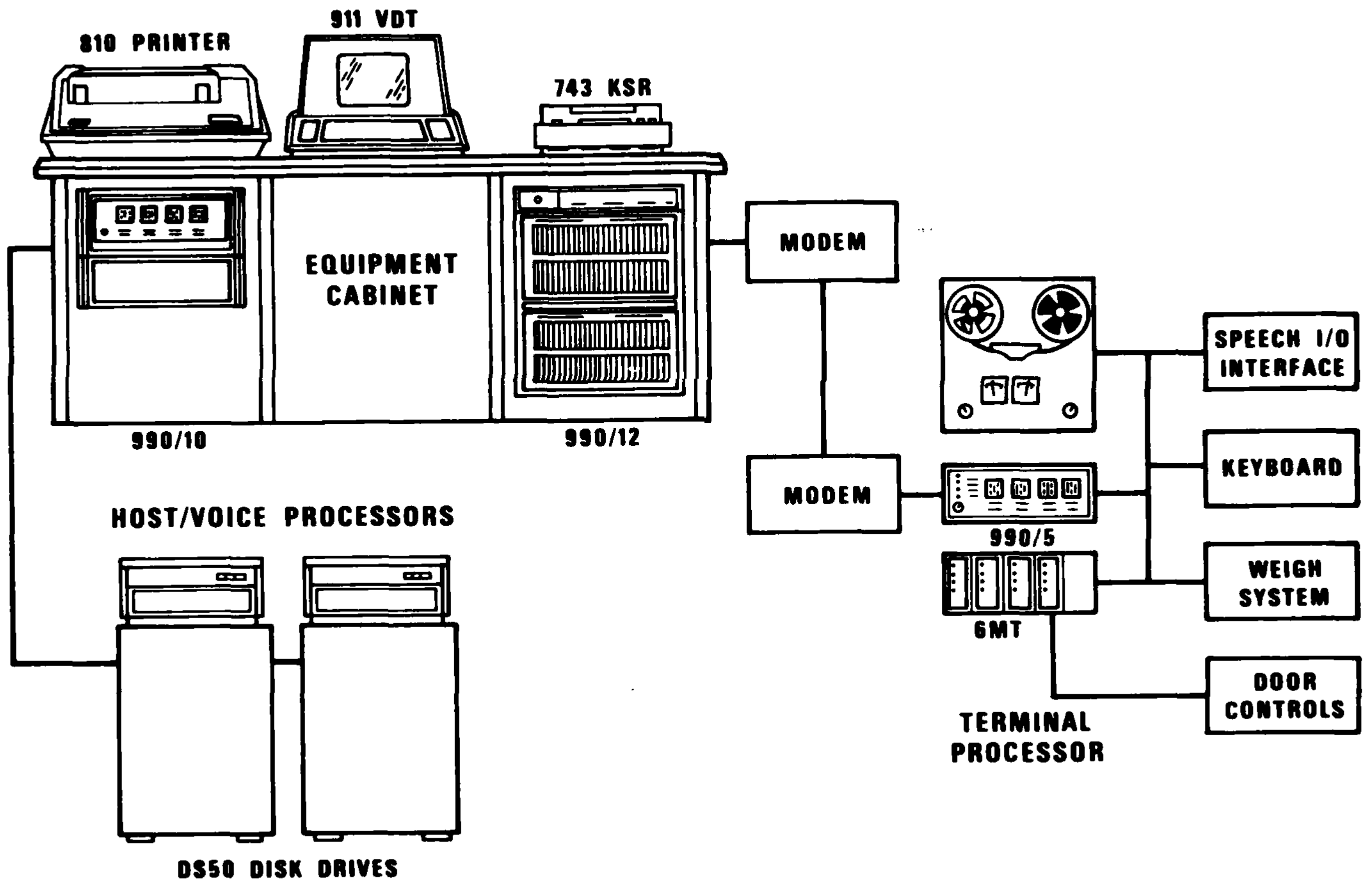


Figure 5 VVU System Components

TABLE 2. VOICE PROCESSOR PARTS LIST

ITEM	QTY
1. 990/12 CPU,17-slot chassis, programmer panel	1
2. Cache memory subsystem (128K bytes)	1
3. 990 communications I/F module (one needed for each terminal processor)	1
4. 9600 baud modem (one needed for each terminal processor)	1
5. Vector comparator (nonstandard part)	1

B. ENTRY CONTROL BOOTH

Because building an entry control booth represented a significant task, several vendors of prefabricated booths were evaluated. A visit was made to two booth vendors who offered integral weigh systems, Campbell Engineering and Mardix Corporation. Both potential vendors were given a technical evaluation. Sites were visited where each vendor had installed booths similar to the booth required by this contract. In addition, both vendors were evaluated for cost and delivery commitment. The following is a summary of the evaluation made.

Both Campbell and Mardix had proven experience in building security booths. Campbell had delivered a small number of booths to Lawrence Livermore Laboratory, and Mardix had delivered a large number of booths to many different customers. Mardix has a product line of security entry systems that includes booths. Campbell has no product line in security entry systems; the booths they have are a side business. Campbell had delivered one (1) booth with an integral weigh system. However, the Campbell weigh system had side loading problems that seem to be inherent in the design approach of weighing the entire booth. Mardix had fabricated a booth for Sandia Corporation that contained a weigh system which weighed only the occupant area thereby minimizing side loading problems.

The Campbell booth was constructed with steel walls and steel structural members. The Mardix booth used special plywood-metal laminated walls that are not easily penetrated but light enough for portability. The framing in the Mardix booth used both steel and aluminum structural members.

The cost of the Campbell booth was estimated to be 50% more than the comparable Mardix booth due primarily to the following reasons:

1. The Campbell booth had a higher material cost than the Mardix booth.
2. The Mardix weigh system had to be modified only slightly, while the Campbell weigh system required a total redesign.

With regard to delivery schedules, Campbell was willing to commit to a 150-day ARO schedule. Mardix committed to a 90-day ARO with a submitted formal quotation. Campbell Engineering did not make a formal commitment.

In conclusion, both vendors met construction requirements, although the Campbell booth was superior in both materials and construction. However, since the Mardix design represented a significantly lower-risk approach with respect to overall cost and delivery, it was chosen over the Campbell booth.

After delivery of the Mardix booth, shown in Figure 6, several additions and modifications were made. The additions included installation of double doors with an integral user terminal (see Figure 7) to separate the equipment bay from the occupant area, installation of equipment racks behind the doors, and installation of a light fixture, dropped ceiling, and sound absorbing material to minimize noise in the occupant area. Figure 8 illustrates the layout of the equipment bay in the entry control booth. In addition, the user terminal, mounted in the right equipment bay door, was lined with sound absorbing material to minimize acoustical reflections. The user terminal has an overall depth into the door of 10 inches, extending 2 inches into the occupant area and recessed 8 inches into the electronics bay. This design was selected to impart a sense of privacy to the user and to keep the microphone from obstructing movement in the booth. The overall height of the user terminal is 45 inches, with the height of the microphone base being 58 inches from the floor of the booth. Due to microphone positions varying from 30 degrees below the horizontal to 60 degrees above the horizontal and to a 10 inch microphone length, the 58 inch height of the base comfortably accommodates from the fifth percentile of female heights through the 95th percentile of male heights [11]. The keyboard is mounted in the lower plate of the terminal with the speaker in an enclosure mounted to the ceiling. Below the terminal is an opening for insertion of a badge to be read via a closed circuit television (CCTV) and to be used in conjunction with an overhead camera for monitoring and backup.

The only modification was to the floor in the weighed occupant area. As supplied, the plywood floor was not sufficiently rigid. Rather than replace the floor with a steel plate (200-300 lbs for an adequate thickness), it was reinforced with angle-iron bracing.

Figure 6 Outside View of Entry Control Booth

Figure 7 View of User Terminal Mounted Inside Booth

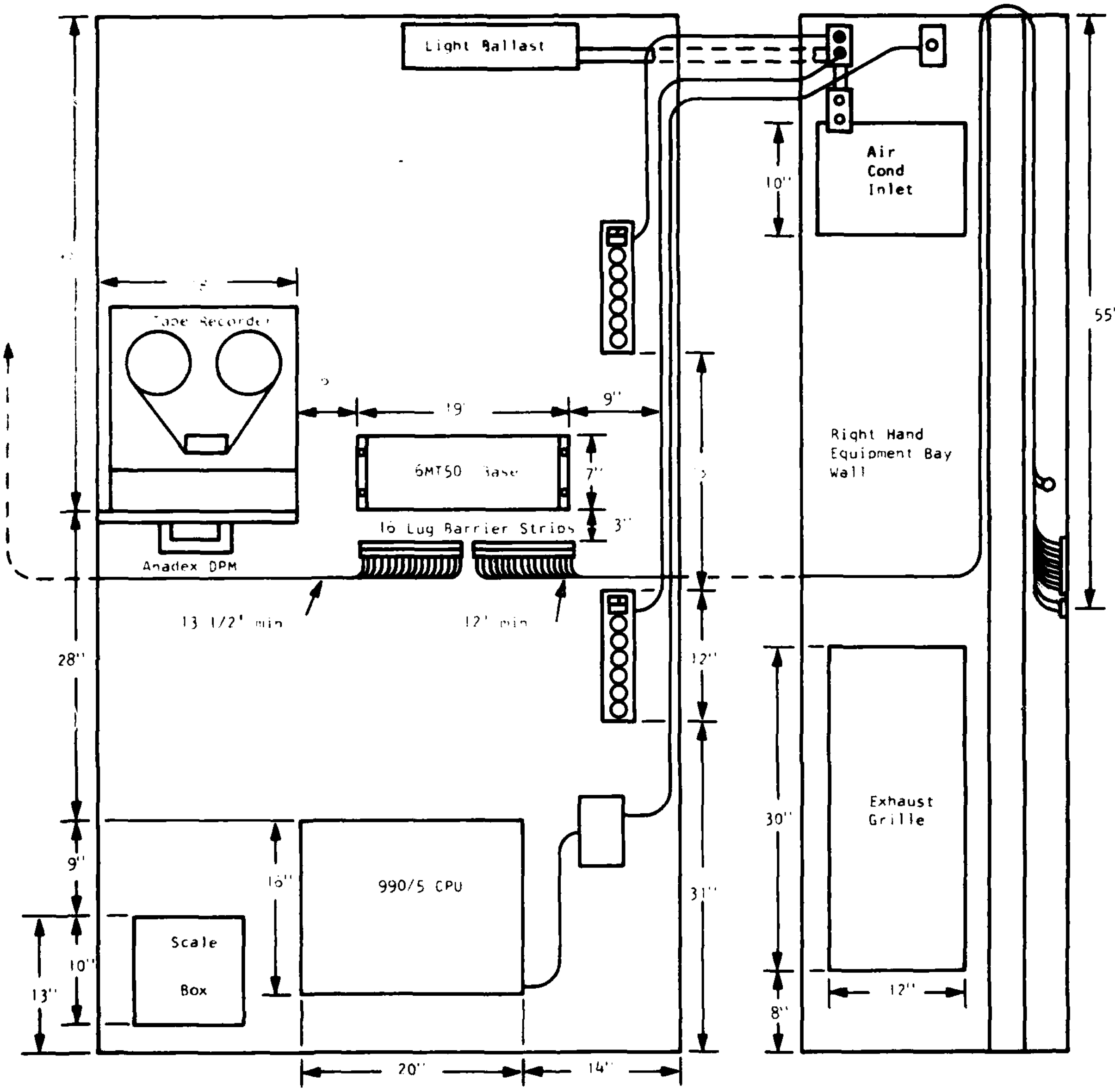


Figure 8 Layout of Equipment Bay in Entry Control Booth

Table 3 is a list of the items used in the entry control booth.

TABLE 3. ENTRY CONTROL BOOTH PARTS LIST

	ITEM	QTY
1.	Mardix VG-300 booth/weigh system	1
2.	Equipment bay doors with integral user terminal	1
3.	BLH Electronics LBPl load cells with a load rating of 250 pounds each	4
4.	Load cell summing box, BLH model 308	1
5.	Anadex indicator, DPM-735 with option B	1
6.	Microphone, Electrovoice 635A	1
7.	Microphone base, plug and shields	1
8.	Speaker	1
9.	Set of 6MT control modules for doors, etc.	1
10.	Terminal Processor	1
11.	9600 Baud Asynchronous Modem	1
12.	24 Vdc Power Supply with integral tester	1

A block diagram of the terminal processor and its components is shown in Figure 9. As can be seen in the diagram the main components are the Speech I/O and Analog Filter boards. These two hardware devices comprise the heart of the system. A combined block diagram of these devices is shown in Figure 10. The terminal processor's CPU interfaces to the Speech I/O board via its communication register unit or CRU, a serial bus. Speech and control data are sent, via the CRU, to a TMC0281 speech synthesizer chip for issuing user prompting phrases. Data for the phrases are stored locally in the terminal processor in volatile random access memory. The Speech I/O board also controls the state of the Analog Filter board by selecting the appropriate input analog filter channel for conversion to digital data and subsequent storage by the terminal processor. The primary responsibility of the Analog Filter board is to amplify and separate the incoming speech waveform into 14 different frequency bands using active passband filters. The speech signal passes through a two stage active passband filter; it is then rectified and passed through an integrating filter section prior to being digitized by the 10-bit analog-to-digital converter residing on the Speech I/O board. Figure 11 shows a typical filter on the Analog Filter board.

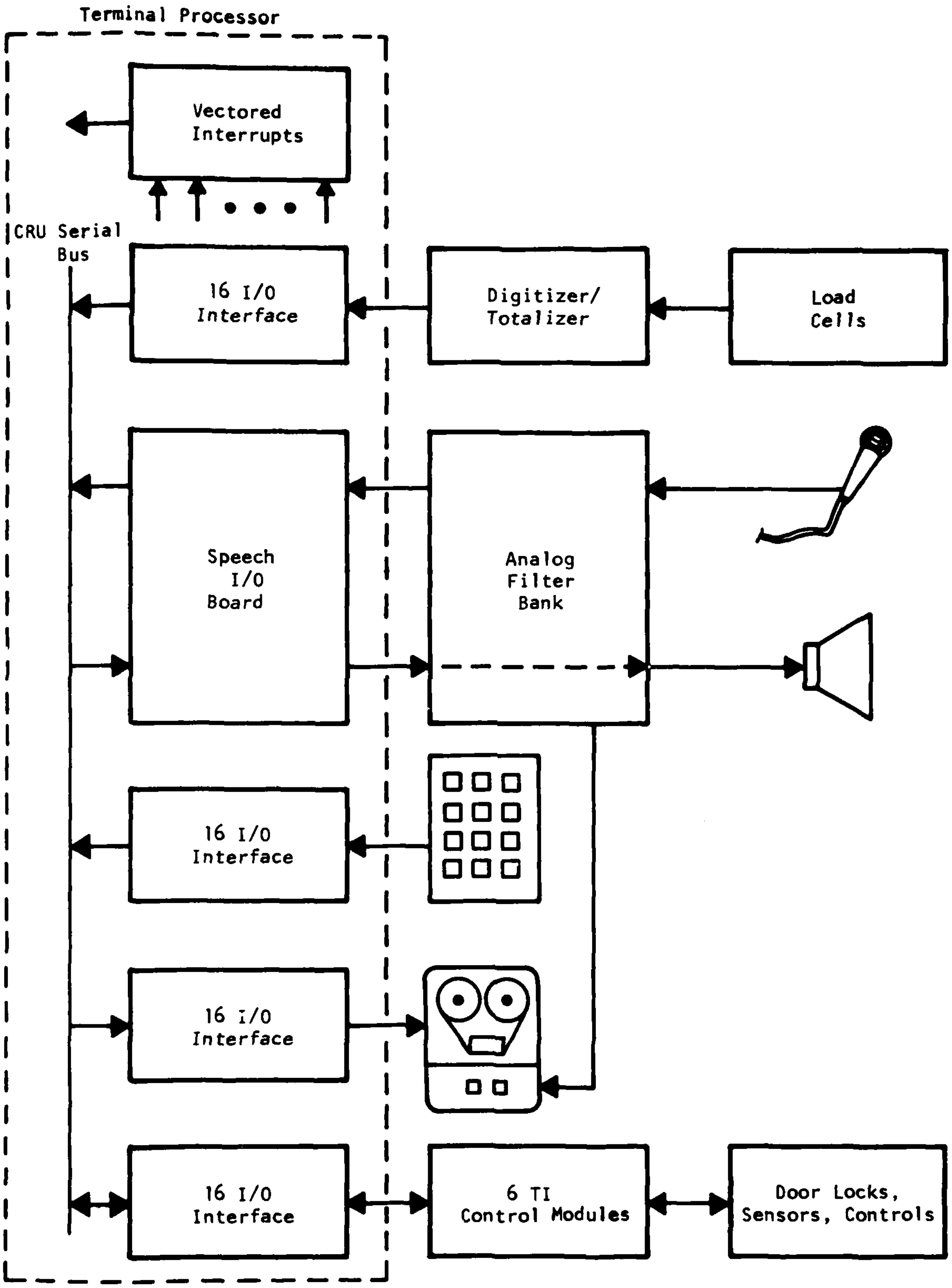
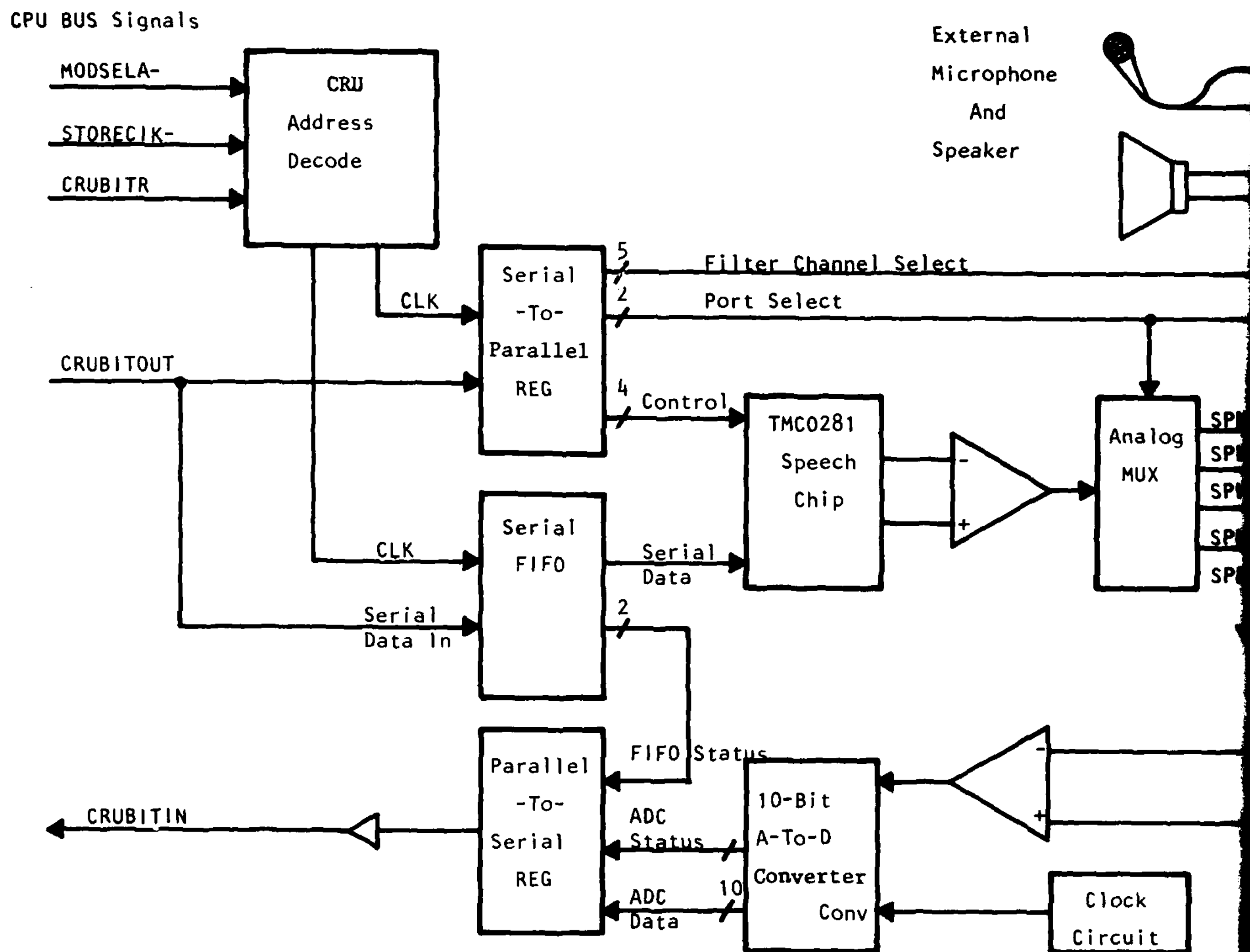
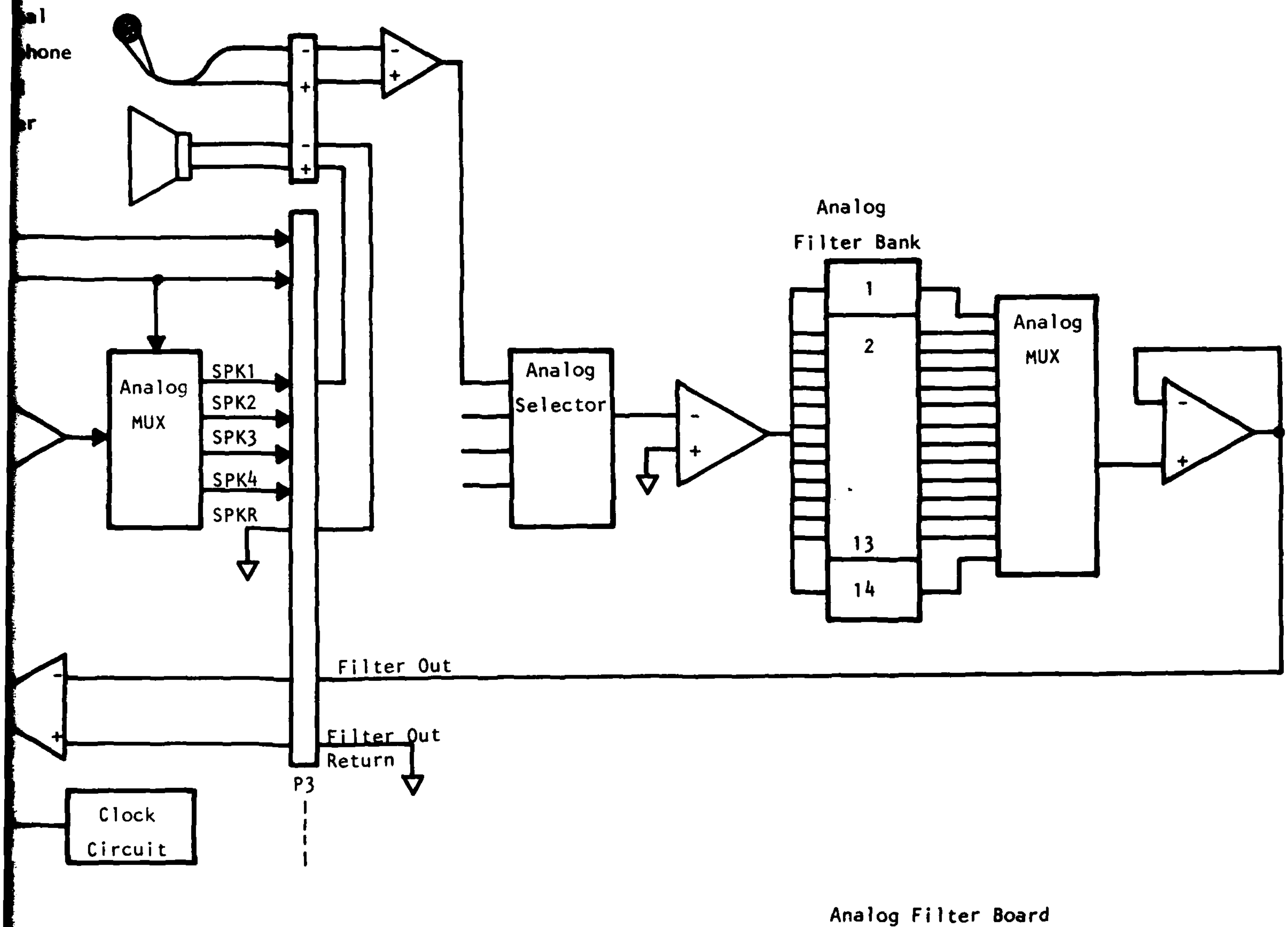


Figure 9 Terminal Processor and Peripheral Interface



Speech I/O Board

Figure 10 Block Diag



ure 10 Block Diagram of Speech I/O System

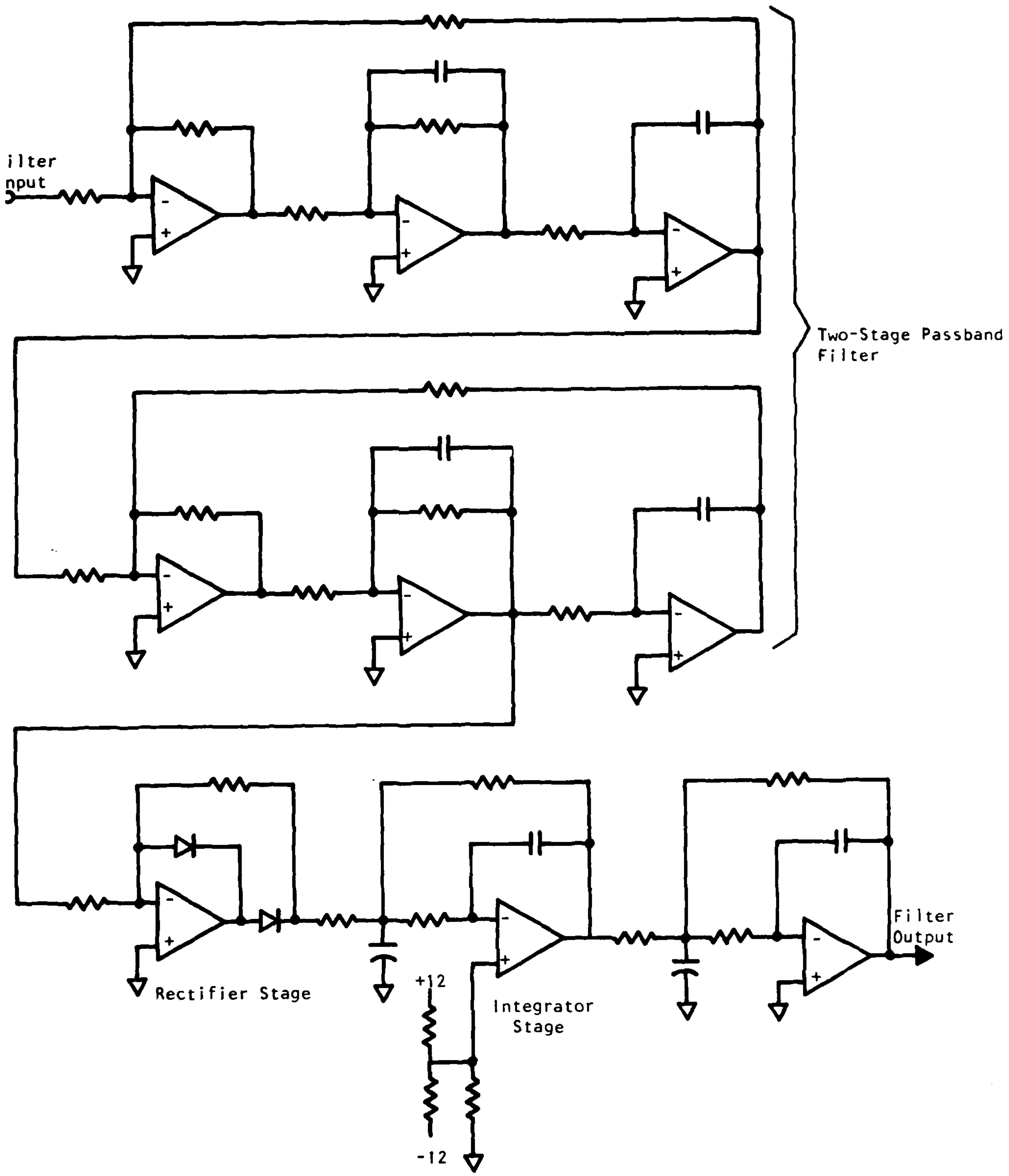


Figure 11 A Typical Analog Filter Board Filter Section

The terminal processor contains the items listed in Table 4.

TABLE 4. TERMINAL PROCESSOR PARTS LIST

	ITEM	QTY
1.	990/5 CPU with 64K bytes, 6-slot chassis, programmer panel, standby power	1
2.	16 I/O TTL data modules (for keyboard, doors and scales)	3
3.	990 communications I/F module	1
4.	12-key keypad	1
5.	Keypad cable	1
6.	Speech I/O to filter board cable	1
7.	990 to 6MT control modules cable	1
8.	Analog filter board (nonstandard part)	1
9.	Speech I/O board (nonstandard part)	1

The terminal processor interfaces to the integral weigh system via its CRU through a standard 16 I/O TTL data module board. The weigh system consists of an ANADEX Transducer Indicator digital panel meter, MODEL DPM-735, a BLH Electronics Model 308 Summing Box, and four BLH Electronics Model LBP1 load cells rated at 250 pounds each. The ANADEX meter provides the excitation voltage for each of the load cells. The load cells return a voltage level based on their deflection to the summing box which adds the returned voltages from each load cell to give a single "summed" voltage. This summed voltage goes to the ANADEX meter for conversion to digital BCD data and ultimate transmission to the terminal processor. Interface to the keyboard in the user's terminal is done with a an additional 16 I/O data module board.

Communication to the voice processor is accomplished via an asynchronous modem which interfaces to the terminal processor through a standard TI 990 communications module board. The terminal processor communicates with the communications board via the CRU.

C. BANDPASS FILTER DESIGN

Figure 12 shows one stage of the filter used on the analog filter board. The following equations can be written from this diagram and can be used to derive the transfer functions of the circuit:

$$\frac{V0 - VA}{R4} + \frac{V2 - VA}{R3} = 0$$

$$\frac{V0}{R0} + V1\left(s * C1 + \frac{1}{R1}\right) = 0$$

$$\frac{V1}{R2} + V2\,(s * C2) = 0$$

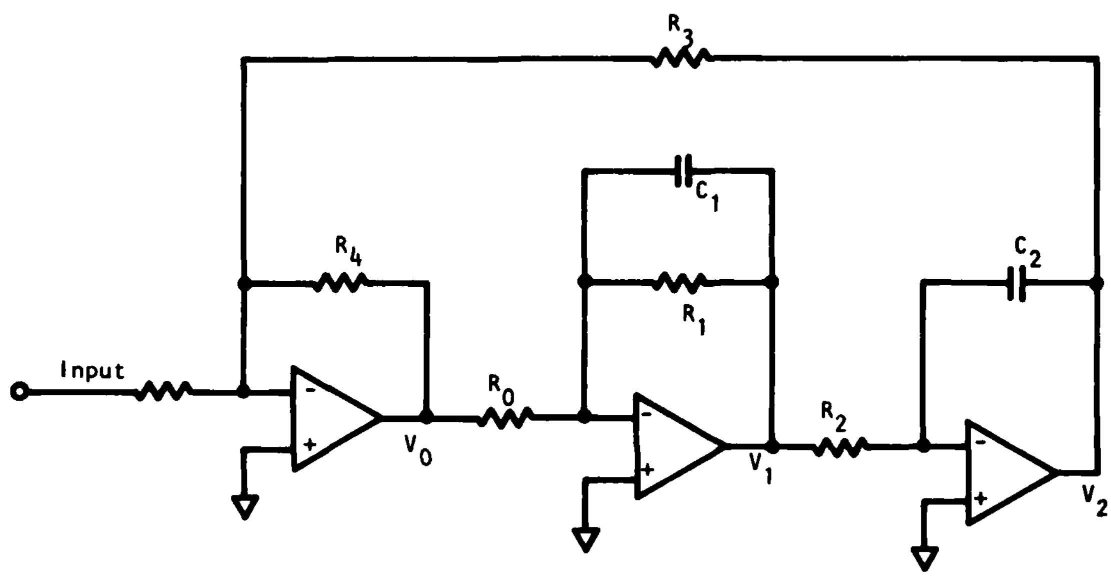


Figure 12 Two Pole Active Bandpass Filter

The transfer functions for V2/VA and for V1/VA are

$$\frac{V1}{VA} = \frac{(1 + R4/R3)}{R0*R2*C1*C2} * \frac{-s*R2*C2}{s*s + s/(R1*C1) + R4/(R3*R0*R2*C1*C2)}$$

$$\frac{V2}{VA} = \frac{(1 + R4/R3)}{R0*R2*C1*C2} * \frac{1}{s*s + s/(R1*C1) + R4/(R3*R0*R2*C1*C2)}$$

By letting

$$CF^2 = R4/(R3*R0*R2*C1*C2)$$

$$BW = 1/(R1*C1)$$

$$G = R4/R3$$

the transfer functions can be written as

$$\frac{V1}{VA} = \frac{(1 + G)}{G} * \frac{-s*R2*C2}{s*s + s*BW + CF*CF}$$

$$\frac{V2}{VA} = \frac{(1 + G)}{G} * \frac{1}{s*s + s*BW + CF*CF}$$

Now, letting $R2*C2 = 1/CF$, $G = 1$, and $C1 = C2 = 10^{-8}$, yields

$R1 = 10^8/BW$, $R0 = R2 = 10^8/CF$ and $R3 = R4$. Hence by selecting the desired center frequencies and bandwidths, the appropriate values for the R's can be determined. The frequency response characteristic for V0, V1, V2 and V1 + V2 for a filter with CF = 350 Hz and BW = 300 Hz is shown in Figure 13.

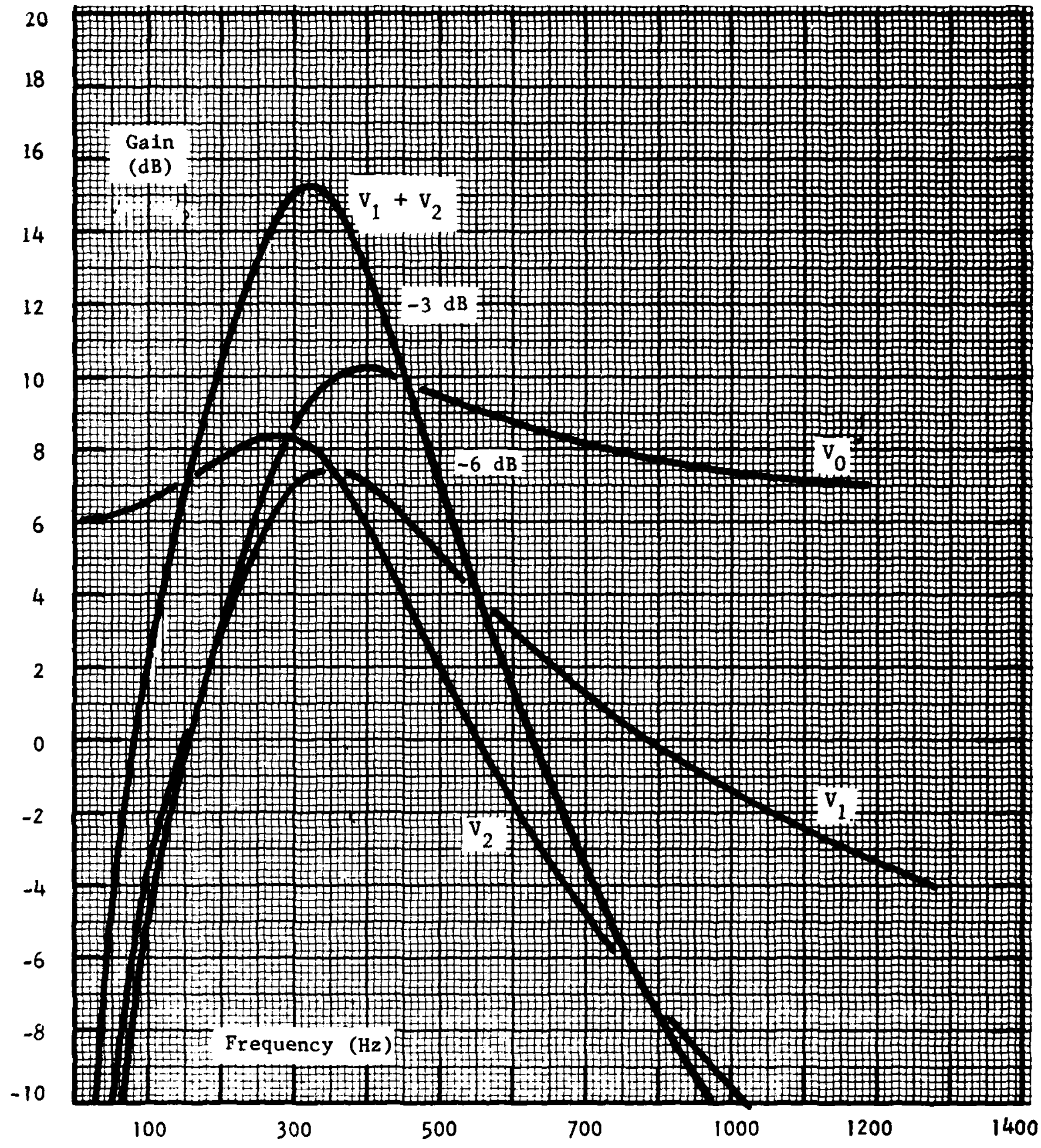


Figure 13 Frequency Response Characteristics for Filter #1.
Nominal CF = 350 Hz
Nominal BW = 300 Hz

SECTION III

SOFTWARE

A. HOST PROCESSOR SOFTWARE

The host is a TI 990/10 minicomputer that executes the DX10 operating system. The function of the host is to maintain the database containing reference data and to provide the operator interface.

The host also provides additional security checks. Each user may be assigned a security level. When a user successfully voice verifies, the host checks that the user has a high enough security level to enter the area. The host also maintains a time record for each user, and users can be assigned permission to enter an area only during assigned times of day.

The host checks for permission to use operator commands from the system console. Only those users assigned permission as operators can issue commands to the system. The operator can issue a wide variety of commands for maintaining, updating, and interrogating the data base, and for controlling the state of the voice and terminal processors. Figures 14 and 15 give two examples of some of the many reports that may be generated by the operator.

The database is stored as a set of key indexed files on the system disk. Voice data and personal data are stored in separate files. The user ID is the primary key used to acquire data from these files.

The operator interface consists of a set of system command procedures, which are stored as text files in three directories on the system disk. The procedures for normal operator interaction are stored in the directory ".F0Y3A5JK", the menus are stored in ".PROC", and the system maintainence commands are stored in ".KM6B34FR". These command procedures prompt for and collect parameters from the operator. The parameters are then passed to tasks which carry out the requests.

Several tasks are normally executing when the voice system is running. These tasks perform communication with the voice processor and database control. The MAPPER task performs message switching functions. It receives messages from the voice processor communications task and sends them to the appropriate task for further processing.

INTASK and OUTTASK perform communications with the voice processor. OUTTASK accepts messages from an intertask communications (ITC) channel (see DX10 manual, volume 3 [11] for an explanation of intertask communications) and sends them to the voice processor. INTASK receives messages from the voice processor, checks them for validity and passes them to MAPPER for further routing.

GETVERFY is the task that accesses the database to obtain reference

data. When a user enters his/her ID on the keyboard in the voice booth, the ID is passed to GETVERFY. GETVERFY reads the reference data for that user and sends the data to the voice processor for use in the verification process.

INTASK, OUTTASK, GETVERFY, and MAPPER are memory resident tasks. The response of these tasks to events is critical to the performance and response time of the host. The tasks are therefore locked into memory at all times so that they can respond immediately to events. The DX10 operating system requires that memory resident tasks be installed in the system program file (.S$PROGA). The other speech related tasks are installed in a program file SVSYS.DATABASE.VOICEX .

1. THE HOST DATA BASE

The data base consists of the following files:

PERSONAL FILE
Contains data relating to nonsecurity attributes of the user such as address, phone number, etc. Only accessed to generate reports.

ACCESS CONTROL FILE
Contains enrollment status of user, console ability, and a copy of the user area control record. This file contains information used to determine whether a user has access to an area. It is accessed after successful voice verification to provide additional security and restrictions such as time of day access and security levels.

AREA CONTROL FILE
Contains a record for each area a user can enter. The record contains the time of day a user may enter an area for each day of the week and an authorization level for that area.

VERIFICATION FILE
Contains voice reference data and verification statistics.

INVENTORY FILE
Contains a record for each user that is in an area. The record contains the area the user is in, the time the user entered the area, and the time user should leave the area.

SYSTEM MAP
Contains status and type information for each portal in the system (only one in the VVU system). Type information includes verification, enrollment and area served.

LEVEL FILE
Contains security level for each area and maximum valid area number.

USER PERFORMANCE SUMMARY

JAN 3, 1980

USER NAME: SMITH, JOHN

USER ID: 000001

USER WEIGHT: 180

ENROLLMENT DATE:..................... JUN 21, 1979

DATE OF LAST SYSTEM ACCESS:.......... NOV 12, 1979

VOICE VERIFICATION STATISTICS

PHRASES FOR SUCCESSFUL VERIFICATION	
1 PHRASE:................................	19
2 PHRASES:...............................	3
3 PHRASES:...............................	0
4 PHRASES:...............................	1
5 PHRASES:...............................	0
6 PHRASES:...............................	0
7 PHRASES:...............................	0
TOTAL SUCCESSFUL VOICE VERIFICATIONS:.....	23
UNSUCCESSFUL VOICE VERIFICATIONS	
NO USER RESPONSE:.........................	5
VOICE MISMATCH:...........................	10
SECURITY VIOLATIONS	
USER/AREA INVENTORY CONFLICT:.............	0
INACTIVE USER:............................	1
UNAUTHORIZED TIME:........................	0
UNAUTHORIZED LEVEL:.......................	0
UNAUTHORIZED AREA:........................	0
SUCCESSFUL ACCESS ATTEMPTS:.................	23
OPERATOR OVERRIDES:.........................	12

NON-VOICE VERIFICATION STATISTICS

SECURITY VIOLATIONS	
USER/AREA INVENTORY CONFLICT:.............	0
INACTIVE USER:............................	0
UNAUTHORIZED TIME:........................	0
UNAUTHORIZED LEVEL:.......................	0
UNAUTHORIZED AREA:........................	0
SUCCESSFUL ACCESS ATTEMPTS:.................	5
OPERATOR OVERRIDES:.........................	6

Figure 14 Sample User Performance Summary

PORTAL PERFORMANCE SUMMARY

DEC 5, 1979

PROCESSOR	PORTAL
1	1

DIRECTION: FROM AREA 0 TO AREA 1	00:00 TO 08:00	08:00 TO 16:00	16:00 TO 24:00	TOTAL
SUCCESSFUL ACCESS ATTEMPTS:........	0	0	0	0
UNSUCCESSFUL ACCESS ATTEMPTS:......	0	0	0	0
USER ID NOT ENROLLED:............	0	0	0	0
USER/AREA INVENTORY CONFLICT:....	0	0	0	0
NO USER RESPONSE:................	0	0	0	0
VOICE MISMATCH:..................	0	0	0	0
SYSTEM MALFUNCTION:..............	0	0	0	0
PRIVILEGE VIOLATION				
INACTIVE USER:.................	0	0	0	0
UNAUTHORIZED TIME:.............	0	0	0	0
UNAUTHORIZED LEVEL:............	0	0	0	0
UNAUTHORIZED AREA:.............	0	0	0	0
OPERATOR OVERRIDES:................	0	0	0	0
PHRASES FOR SUCCESSFUL VERIFICATION				
1 PHRASE:........................	0	0	0	0
2 PHRASES:.......................	0	0	0	0
3 PHRASES:.......................	0	0	0	0
4 PHRASES:.......................	0	0	0	0
5 PHRASES:.......................	0	0	0	0
6 PHRASES:.......................	0	0	0	0
7 PHRASES:.......................	0	0	0	0

DIRECTION: FROM AREA 1 TO AREA 0	00:00 TO 08:00	08:00 TO 16:00	16:00 TO 24:00	TOTAL
SUCCESSFUL ACCESS ATTEMPTS:........	0	0	0	0
UNSUCCESSFUL ACCESS ATTEMPTS:......	0	0	0	0
USER ID NOT ENROLLED:............	0	0	0	0
USER/AREA INVENTORY CONFLICT:....	0	0	0	0
NO USER RESPONSE:................	0	0	0	0
VOICE MISMATCH:..................	0	0	0	0
SYSTEM MALFUNCTION:..............	0	0	0	0
PRIVILEGE VIOLATION				
INACTIVE USER:.................	0	0	0	0
UNAUTHORIZED TIME:.............	0	0	0	0
UNAUTHORIZED LEVEL:............	0	0	0	0
UNAUTHORIZED AREA:.............	0	0	0	0
OPERATOR OVERRIDES:................	0	0	0	0
PHRASES FOR SUCCESSFUL VERIFICATION				
1 PHRASE:........................	0	0	0	0
2 PHRASES:.......................	0	0	0	0
3 PHRASES:.......................	0	0	0	0
4 PHRASES:.......................	0	0	0	0
5 PHRASES:.......................	0	0	0	0
6 PHRASES:.......................	0	0	0	0
7 PHRASES:.......................	0	0	0	0

Figure 15 Sample Portal Performance Summary

OPTIONS FILE
Contains option enable flags for system behavior.

SPOOL FILE
Contains system event history similar to a system log.

2. THE HOST TASKS

The following tasks are provided in the host:

2.1 DATA BASE MAINTENANCE TASKS

AREACTRL	Maintain area control records
CHGINVT	Manual change of inventory
CLOSE	Close KIF file
CLRDIAG	Clear diagnostic mode
CLRDISK	Clear research file
CLRRSCH	Clear research mode
COMMENT	Send a comment to the system log
EMITIE2	Emit init enroll command
EMITPUR	Emit process user request
ENROLITC	Send operator enrollment response to ENROLLM
ENROLLM	Perform host function of enrollment
GETVERFY	Perform host function verification
HOSTAC	MSGOPER messages from host
LOGMSG	LOGSPOOL messages
LOGSPOOL	Log system activity to printer and circular file
MSGOPER	Prints alert messages and waits for responses
ODDJOB	Abort enrollment, zero's user performance statistics
OVERRIDE	Sends override to VP; checks inventory for violations
PDACCESS	Creates access control report
PDAREA	Creates users in an area report
PDLOG	Creates system log report
PDMAP	Creates system map report
PDPERSNL	Creates personal data report
PDPORTAL	Creates portal performance report
PDVERFY	Creates user performance report
PRNTDISK	Creates research data report
PRNTREF	Creates reference data report
RSCHDISK	Writes research mode file
SETDIAG	Set diagnostic mode
SETOP	Sets system option record
SETRSCH	Sets research mode
STOPBELL	Sends stop bell command to MSGOPER
TERMAC	MSGOPER messages from terminal
TESTLOG	Tests LOGSPOOL
TESTRSCH	Test host end of research mode
TSTMSGOP	Test MSGOPER
UDACCES1	Enter and modify access control file
UDACCES2	Deletes, changes user state for access
UDAREA	Maintains area inventories
UDMAP	Modifies system map

UDPORTAL	Maintains portal performance statistics
UDPRSNL1	Enter and modify personal data
UDPRSNL2	Modifies the data base enrollment status when a user is reenrolled or deleted
USERSUM	Creates enrolled user summary
USRSTAT	Creates user's verification statistics report
ZEROPORT	Zero's portal performance statistics

2.2 OVERHEAD TASKS

ASGNLUNO	Assigns lunos
ASSIGN	Assigns lunos from a file
DIAGDISK	Writes diagnostic mode file
DOWNTOAD	Downloads VP object
DUMPCHAN	Displays ITC channels
INDSR	Input DSR
INTASK	Input driver for VP
HEART	Sends and monitors heartbeat messages
HOSTUPDN	Sends host up/down messages
INPDT	PDT for INDSR
LOOPBACK	Sends and receives loopback messages to a VP
MAPPER	Controls tasks and replicates ITC messages
OUTDSR	Output DSR
OUTASK	Output driver for VP
OUTPDT	PDT for OUTDSR
PURGE	Remove ITC messages from a channel
QUEDSR	Input queue DSR
QUEPDT	PDT for QUEPDT
SYSERROR	Sends system error message to MSGOPER
TIMER	Receives an ITC message, holds it, then resends it

2.3 SECURITY TASKS

AUTHVIOL	Generates users in authorization violation report
CATCH	Periodically checks inventory for users past time
CHEKINV	Checks a user for inventory violations
CHEKSECR	Sends request to perform a command to SECMONT
INITLEVL	Loads area levels
LOGIN	Accepts password
MTHDAY	Converts julian day to month and day of month form
PDLEVEL	Creates area level report
REWIND	Rewind report temp files
SECMONT	Checks security for verifications and commands
SETSYN	Sets synonym to proc directory
STATUS	Generates status of user report
TESTSECR	Tests SECMONT
UDCONS	Updates console privileges
UDLEVEL	Maintains area level record

2.4 UTILITY PROCEDURES USED BY VOICE TASKS

ABORT	Abort program
CALDATE	Julian to calendar date
CONMSG	Sends alert message to MSGOPER
DELTCUR	Delete current KIF record
DELTKEY	Delete a KIF record
ENTMSG	Enrollment terminated message
GETDAY	Time of day and day of week
GETIME	Get time and day
GETITC	Read an ITC channel
GETNAME	Get the name corresponding to an ID
IARMSG	Invalid access request message
IIDMSG	Invalid ID message
INITMAP	Load system map
INITOP	Load system options
INSRTR	Insert a KIF record
IOSVC	General I/O SVC
JULDATE	Calendar date to julian date
JULDAY	Calendar date to julian date
LOADMAP	Replace system map
LOADOP	Replace system options
LOGMSG	Sends log message to LOGSPOOL
OPEN	Open KIF file
OPENCALL	Open a file and return characteristics
OVRMSG	Sends override message to VP
PKEYREAD	Read KIF record by primary key
PRTMSG	Update portal statistics message
RD$VDT	Read VDT
PUTITC	Send an ITC message
READCUR	Read current KIF record
READEQG	Read KIF record >=
READKEY	Read KIF record by key
READNEXT	Read next KIF record
READPREV	Read previous KIF record
RERITR	Rewrite KIF record
SENDITC	Sends some ITC messages
SCI$IF	Interface to SCI
SETEQG	Set currency to scan key indexed file
SVCIO	General I/O SVC
TMEMSG	Message to timer
UNLOCK	Unlock KIF record
WR$VDT	Write to VDT

2.5 MAIN TASKS

MAPPER
Controls task scheduling of voice tasks, relays ITC messages to tasks, and provides debugging information. Almost all ITC messages sent in the system will pass through MAPPER. GETVERFY and ENROLLM tasks send and receive large amounts of reference data to the voice processor. To make this process more efficient, they

communicate with INTASK and OUTTASK by means of shared memory. A message is sent through MAPPER to notify the destination task that data are present in shared memory. The data are transmitted directly to the voice processor, avoiding the system overhead involved in intertask communication.

GETVERFY
Maintains verification database, and monitors and controls the verification process. Hands control to ENROLLM in cases where an enrollment might be required.

3. HOST SCENARIOS:

3.1 VERIFICATION

1. A process user request is received by an input DSR and sent to MAPPER.

2. MAPPER sends the request to SECMONT and GETVERFY and then activates them.

3. SECMONT makes all necessary security checks on the user and sends a command to GETVERFY indicating the result. If the result is a security rejection, SECMONT sends a message to LOGSPOOL.

4. GETVERFY reads the user's verification file and waits for the command from SECMONT. When it arrives, GETVERFY sends a response to the voice processor. If voice verification is required, the reference data is sent to the processor with the response. A message is sent to UDPORTAL to update the portal statistics.

5. When the verification is complete, the result and update data are sent to GETVERFY. GETVERFY updates the statistics, rewrites the verification record, and sends a message to MSGOPER.

6. When the door is opened by the user, the door-opened message is sent to MAPPER who sends it to GETVERFY and UDAREA. UDAREA updates the inventory and GETVERFY makes certain all users at that portal have been accounted for and sends the door-opened to LOGSPOOL.

7. LOGSPOOL receives messages indicating security check failure, verification result, and a door being opened.

3.2 ENROLLMENT

1. A process user request is received by an input DSR and sent to MAPPER.

2. MAPPER sends the request to SECMONT and GETVERFY and then activates them.

3. SECMONT will reject the user because he will not be active and will send the result to GETVERFY.

4. GETVERFY will try to read his verification record and will find that the record does not exist. When the command from SECMONT arrives, ENROLLM will be sent an enrollment check command from GETVERFY. The command will include an address of a verification record to use for the new user.

5. When the enrollment check command passes through MAPPER, it will bid ENROLLM and pass it the message.

6. ENROLLM will see if an enrollment window has been opened for this user. If it has not, the enrollment is stopped and a message sent to MSGOPER.

7. If the window is open, an alert message is sent to MSGOPER and ENROLLM waits for a response from the operator.

8. If the response is no, the enrollment is stopped and a message is sent to LOGSPOOL.

9. If the response is yes, the enrollment command is sent to the portal and ENROLLM times out the enrollment process.

10. When the enrollment is finished, ENROLLM initializes a new verification record and sends a message to GETVERFY indicating completion of enrollment.

3.3 REQUEST TO PERFORM A COMMAND

1. In the proc of each command is a call to CHEKSECR.

2. CHEKSECR reads the password and ID of the operator; the ID of the user, if any; and the command number and set.

3. CHEKSECR sends the information to SECMONT which makes the authorization check. SECMONT sends the information about the command and its decision to LOGSPOOL. SECMONT sends a response to CHEKSECR which assigns that value to a synonym. The proc stops SCI if the request is not allowed.

B. VOICE PROCESSOR SOFTWARE

The voice processor consists of a TI 990/12 minicomputer with 128 Kbytes of memory and communication interfaces to both the host and terminal processors. The function of the voice processor is to perform, under control of the host processor, verification and enrollment of in-

dividuals using preprocessed voice data, as sent from the terminal processor via a 9600 baud asynchronous serial communications channel. Although the use of the 990/12 as the voice processor and the size of its memory were selected to provide the processing power for servicing multiple terminal processors (and memory size for servicing two terminal processors), the software installed on the delivered voice processor has not been expanded to provide for more than one terminal processor.

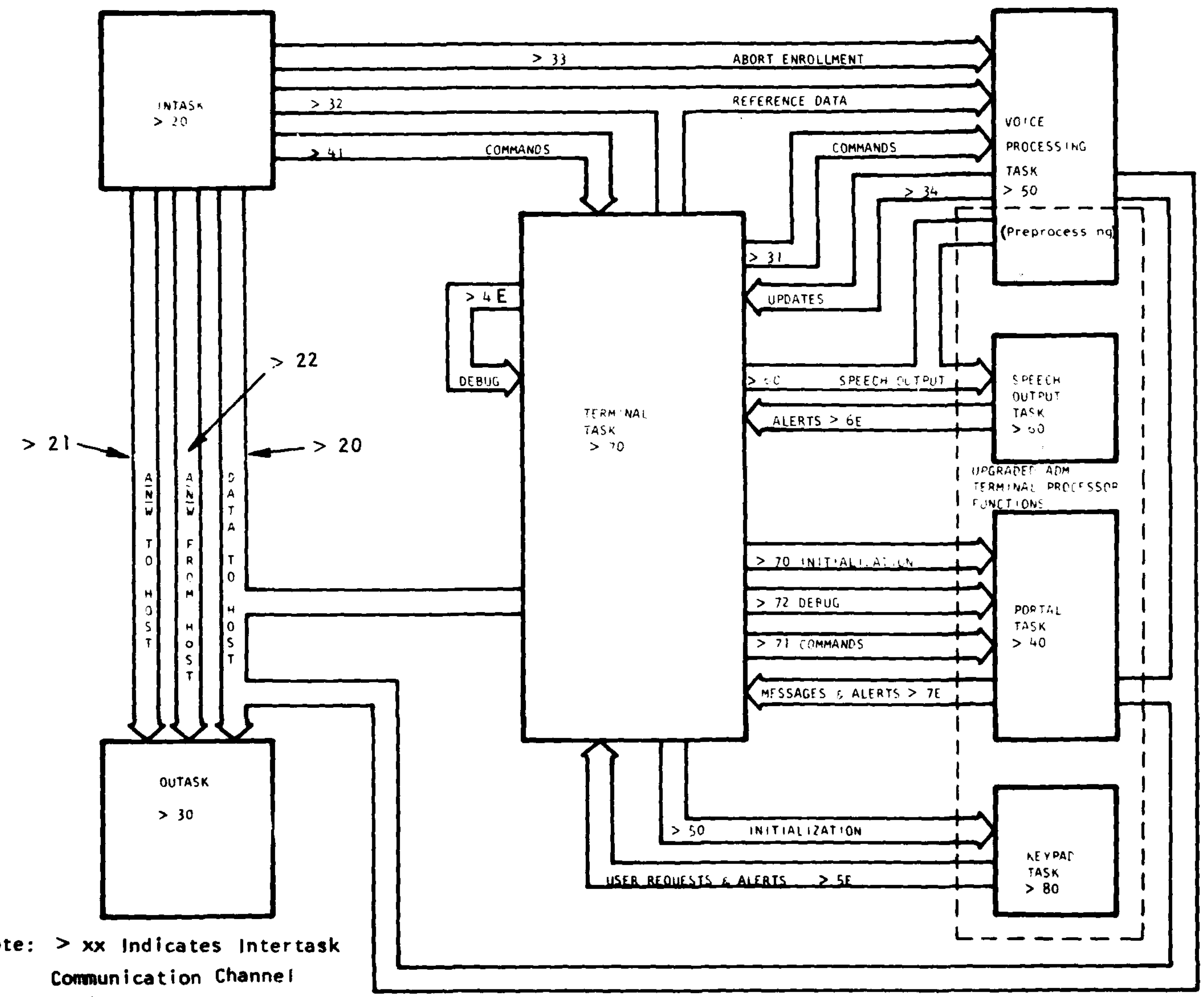
The voice processor functions are performed by a set of cooperating, asynchronous tasks, primarily communicating with each other via dedicated "intertask communication (ITC)" channels (dedicated memory areas). Each task is allocated a portion of CPU time (50 ms time-slice) by a priority scheduling algorithm which accounts for interrupts, suspended tasks, etc. This task scheduling is part of the underlying "TX 0" operating system that also provides memory management, interrupt processing, intertask communication, interval timing, task initialization, etc. TX990 is described in more detail in the "TX990 Operating System Documentation." [12]

The interrelationship of the asynchronous tasks on the voice processor is shown in Figure 16. For comparison, the task block diagram for the DSG voice processor is shown in Figure 17, where the functions within the dashed line have been moved to the terminal processor for the VVU system to reduce the processing necessary for each voice terminal, enabling one voice processor to service many lower cost terminal processors.

Almost all the tasks have been written exclusively in Pascal. However, a few selected procedures have been written in 990 assembly language as needed for speed. All the TX990 operating system was written in 990 assembly language.

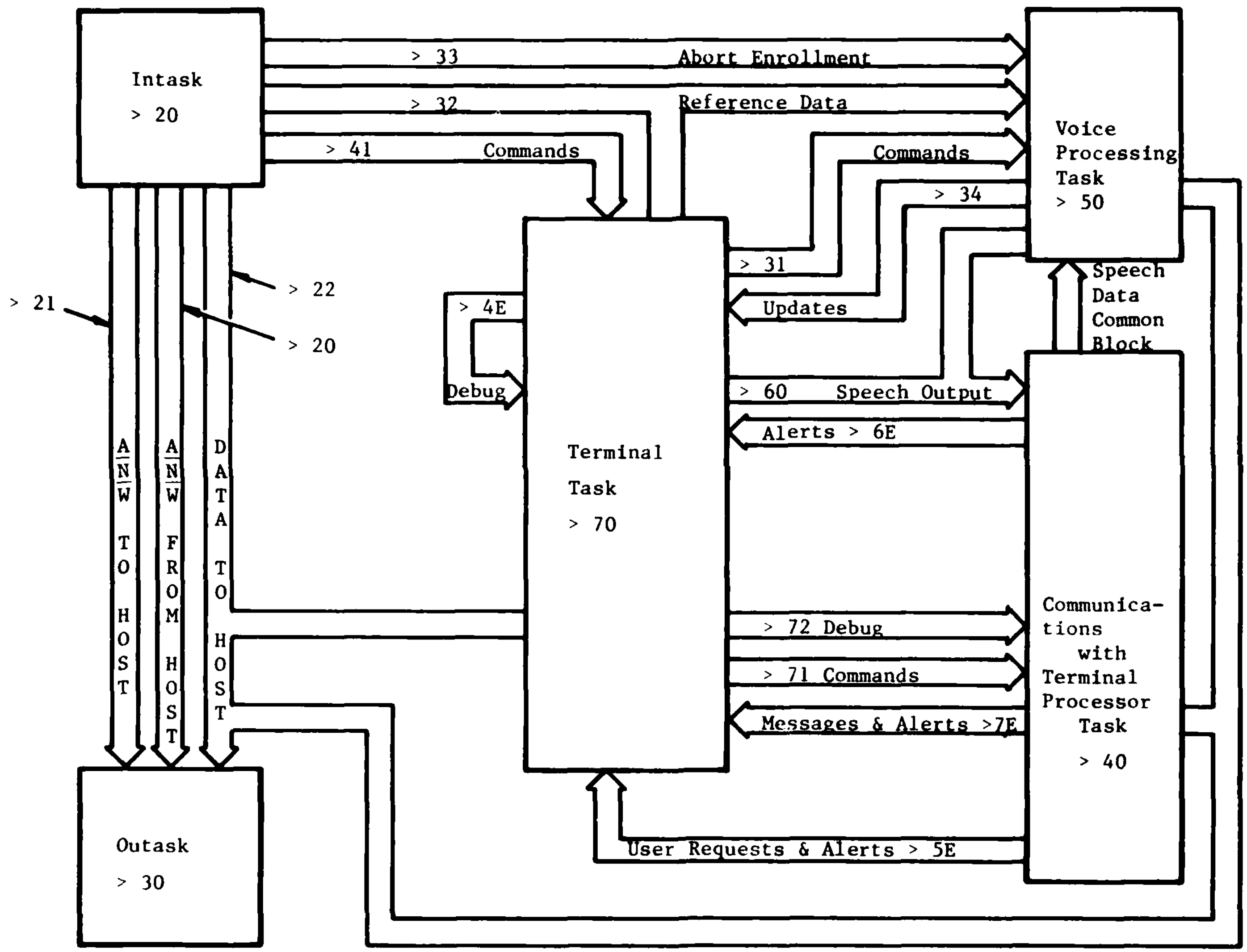
1. TASK TO COMMUNICATE WITH THE TERMINAL PROCESSOR

The communications routines at both the voice processor and the terminal processor ends are designed to accommodate errors in transmission by requiring positive acknowledgments (ACKs) to be received by the sender before discarding any message that has been sent. Although it is doubtful that errors will ever be introduced by the channel itself (at least in our environment), errors can occur (missed characters) due either to full buffers in the receiver or to missed interrupts because other higher priority, noninterruptable processing is occurring that lasts longer than the time between the receipt of two sequential input characters (approximately every 1 ms). Any time such an error occurs, a retransmission is required, which increases the load on the communication channel. In addition to the loss of the original message, the error can also be due to the loss of the return ACK. This loss of ACKs means that the sender, after an appropriate waiting time, must spontaneously (and needlessly) retransmit the block of data corresponding to the missing ACK. The receiver, of course, will have to discard the extra message, since it had already been received. The object then is to reduce the maximum possible delay between the receipt of input characters as much as possible in order to decrease the frequency of characters



Note: > xx Indicates Intertask Communication Channel Numbers

Figure 16 Inter-Task Communications for DSG's Voice Processor



Note: > xx Indicates Intertask
Communication Channel Numbers

Figure 17 Inter-task Communications for the VVU System's Voice Processor

(and hence messages) being missed.

The task concerned with communications with the terminal processor operates using this error-recoverable protocol. Data are received from the terminal processor on an interrupt driven basis with each byte being placed in a circular input queue (200 bytes long) of numbered input messages, by an input device service routine under control of the operating system. After an entire block of data is received correctly, an acknowledgment is sent to the terminal processor. If messages prior to a correctly received message are missing, NAKs are also sent to flag the missing messages. As time permits, another device service routine removes data from the input queue, checking parity, stripping off synchronization and "transparent" characters, and blocking the data for transmission to another task either via ITC channels or via a common block, in the case of preprocessed speech data. One exception is that "heartbeat" messages received from the terminal processor are acknowledged back to the terminal processor, but are not passed on.

Data are also received from other voice processor tasks via ITC channels, and are blocked for transmission to the terminal processor, with appropriate parity and synchronization character generation and message block number assignment. After this processing, the data are queued for transmission to the terminal processor (along with ACKs and NAKs to the incoming messages) via an output device service routine, also operating on an interrupt driven basis under control of the operating system.

In summary, the terminal processor communications task provides a transparant interface between the voice processor tasks and the tasks that have been moved from DSG's voice processor to the terminal processor in the VVU system. The software at the terminal processor end of this communication link is described in Part C.2 of this section.

2. TERMINAL TASK

The name for this task was assigned before the inception of the "terminal processor" in the VVU system and, regrettably, imposes additional confusion on the description of the software system. Rather than having any direct relation to the terminal processor, this task controls the state of the voice processor and its control of the terminal processor. In fact, the terminal task is primarily a controller for a large state table, driven both by commands from the host processor and by requests and responses from the voice processing task and the tasks in the terminal processor.

When the booth is empty, the terminal task is idle, except for responding to occasional heartbeat messages from the host. (Terminal processor heartbeats are handled entirely in the communications routine and do not affect the terminal task.) When a door to the booth is opened, the terminal task starts a timer which will prompt a message to close the door if the door is held open too long (one minute), or which will disable the booth after three such messages have been prompted. If the door is closed without any weight in the booth, the terminal task re-

turns to its idle state. However, if weight is present after the door closes, commands to lock the booth doors are sent to the terminal processor (at least one booth door is always locked), the initial timers are reset, and additional timers are initiated to check that some user operation is requested (by the input of a valid identification number via the keyboard) within some prescribed time interval to help insure that the user is not sabotaging the booth. Once a valid identification number has been input, the number is received by the terminal task and is passed on to the host for determination of the proper action to be taken. If the user is to be either enrolled or verified, the request is received by the terminal task, causing a state transition in its state table, and the request is passed on to the voice processing task in the voice processor, which initiates the appropriate prompting and processing of speech input. In addition, the terminal task will also initiate still more timers to insure that the user responds within a prescribed interval and that the entire verification or enrollment does not take an abnormally long time.

If no valid input is received from the keyboard in the booth within a prescribed time, both doors remain locked, a terminal task state table transition is made, the user is prompted to "call for assistance," and exit from the booth is possible only upon receipt of an operator initiated override from the host (security control).

Once a user has been verified, the terminal task relates the success to the host, undergoes a state transition, and if the weight remaining in the booth after the verified user's weight is subtracted exceeds forty pounds, the timers waiting for keyboard entry are initiated again and the above procedure is repeated (with appropriate additional state transitions) for another user identification. Also, if a user is not verified, the above procedure is repeated (users may try to verify as many times as they wish). Once either the remaining weight is less than forty pounds (for a verification) or an enrollment has been completed, another terminal task state transition occurs, one of the doors is opened (opposite from entry for verification; same as entry for enrollment) and the door-open timers are again initiated.

In addition to these monitors during normal operation, the terminal task will also notify the host of abnormal conditions existing in the booth, such as both doors being opened at the same time, in which case portal security has been lost. In such cases of monitor abnormalities, the terminal task will send a message to the host which will be printed on the alert terminal in security control to prompt appropriate action by security.

3. VOICE PROCESSING

Another unfortunate semantic confusion exists with the naming of the voice processing task, which is one of the tasks running on the voice processor. Nevertheless, the function of the voice processing task is to either verify or enroll a single user by command from the host, as passed on by the terminal task. Hence, the voice processor task sits idle until a command is received via an ITC channel from the

terminal task.

If a received command is to verify a user, the task first selects the set of four phrases to be prompted based upon both a random number generator and the frequency of usage of the words, as stored in the reference file for the user as sent from the host processor. (Readers desiring more details on the algorithm are referred Section IV and Appendix I.) The voice processor then transmits a message to the terminal processor to prompt the first selected phrase, to collect input speech data and to return the data to the voice processor. The communications task in the voice processor relays the data to the voice processing task via a common block area of memory (to avoid extra movement of this large bulk of data) as it is received from the terminal processor, and if the reference data for all of the prompted words have arrived from the host, the comparison between the input and reference data is begun. As soon as a match has been made between the input and the reference, an abort signal is sent to the terminal processor to stop collecting and processing data. In case the match is not sufficiently good, and the algorithm requires more speech data, another message is sent to the terminal processor to prompt another phrase, and this process is repeated for up to seven phrases total. Whenever a final decision is made to verify or not-verify the speaker, the decision is sent to the terminal task, which uses the decision to make the appropriate state transition in its state table, thus prompting further action (another verification attempt or the user exiting the booth). If the user is verified, the user's reference data are updated, and this updated reference file is sent back to the host for storage on disc.

Whenever the terminal task passes the identification number of a user that is not enrolled back to the host, if the personnel and security information for that number have already been properly entered and security has specifically allowed the user's enrollment at this time, the host will send an enrollment command to the terminal task in the voice processor, which in turn is relayed to the voice processing task. Upon receipt of an enrollment command, the voice processing task begins prompting the first of at least twenty phrases used to enroll the user. At least the first four phrases are used to establish initial reference patterns, which are then used to compare with subsequent input data. The largest part of these subsequent comparisons is identical with the processing used during a verification, and hence uses the same procedures. If enrollment is completed successfully, the generated reference file is sent to the host, along with the user's identification number and weight for storage on disc. If enrollment is terminated unsuccessfully, no reference data are returned, but a message is sent to the host telling it of the termination. If during enrollment, security were to determine that the user was not speaking correctly by monitoring his speech, the operator could abort the enrollment from the command terminal on the host. In this case, a message is received at the voice processing task directly from the host communications task (INTASK) and the enrollment is aborted. The terminal task does not know (or care) whether the abort came from the host or was generated internally by the voice processing task due to inadequate data. The list of reasons why enrollment may be terminated by the voice processing task itself is given in

Section IV.

After enrollment has either completed or terminated, the appropriate messages are sent to the terminal processor via an ITC message to the terminal communications task, which will inform the user of the completion or termination, will prompt him to return via the door he entered, and will direct the booth hardware to unlock the door.

4. HOST COMMUNICATIONS TASKS

Communications to and from the host are controlled by two separate tasks: INTASK and OUTTASK. The actual device drivers in the operating system are almost identical to those used for communicating with the Terminal Processor.

During its time slice, INTASK looks for an input message to be processed. If none exist, the task returns the remainder of its time slice to the operating system. If messages exist, INTASK processes them until either its time or the queue of incoming messages is exhausted. If the incoming message is protocol (ACK, NAK, WAK), it is passed on to OUTTASK for appropriate action. If it is text, INTASK passes it on to the appropriate ITC channel to be sent to another task, and passes an ACK to OUTTASK for return to the host. If the ITC channel is full, a WAK is returned to the host through OUTTASK. If either INTASK's input queue is full or a transmission error occurs for a recognizable message, a NAK is returned to the host. All of these appropriate actions are implemented with a state table similar to that described in the Terminal Task section.

OUTTASK is responsible for messages being transmitted from the voice processor to the host. The procedure GETMSG places messages from ITC channels into a queue of areas acquired from the heap, until there are no more messages, the maximum number (32) of messages awaiting transmission has been reached, or the heap has been exhausted. Initiation of the next message to be transmitted is done by SNDMSG, which also preloads counters for TIMMSG to decrement for determining when to spontaneously retransmit unacknowledged messages. After three retransmissions, unacknowledged messages are discarded. (Note that the error recovery capability of host-to-voice processor communications is not nearly as sophisticated as the voice-to-terminal processor communications, which never discards unacknowledged messages.) The procedure GETFROM uses ACKs, NAKs and WAKs from the host (via an ITC from INTASK) either to initiate retransmissions (for NAKs and WAKs) or to remove messages from the queue of outgoing messages (for ACKs). The procedure GETFOR queues ACK, NAK and WAK messages received via an ITC channel from INTASK for transmission to the host in response to messages received at the voice processor from the host.

C. TERMINAL PROCESSOR SOFTWARE

The terminal processor consists of a TI 990/5 minicomputer with associated hardware. The function of the terminal processor is to perform

local processing at the voice booth, which consists of speech synthesis, data collection, speech preprocessing and portal control. The terminal processor communicates with the rest of the voice authentication system via a serial communications interface. It receives instructions from the voice processor and passes speech and event data back to the remainder of the system.

Terminal processor functions are performed by a set of cooperating asynchronous processes. Each basic function in the system is performed by a separate process. These processes communicate with one another through semaphores and buffers in shared memory. Each process is allocated a portion of CPU time by a priority scheduling algorithm.

1. OPERATING SYSTEM

The terminal processor runs the TIPMX operating system. This operating system provides the kernel functions of task scheduling and synchronization. TIPMX is described in detail in the "TI PASCAL MICROPROCESSOR EXECUTIVE USER'S GUIDE". [13]

2. SPEECH PROCESSING

Several speech processing tasks are performed in the terminal processor. First, the terminal processor is responsible for speech synthesis for prompts and messages. The data for the speech synthesizer is stored locally in the terminal processor's memory. Synthesis is initiated via a message from the voice processor using a message identification code. The terminal processor uses this code to find the synthesis data, then causing the message to be spoken.

Speech data collection is initiated by a prompt. The identity of the four words in the prompt is provided in the request from the voice processor. The terminal processor starts the speech synthesizer and begins collecting data (at 10 millisecond intervals). Data are discarded until twenty samples before the end of the prompt, at which time "preprocessing" of the data begins. Preprocessing first regresses the data using sine/cosine basis functions. The data are then normalized and scaled, and the result is blocked for transmission to the voice processor. After each group of 8 samples has been collected and processed, the block is transmitted to the voice processor over the serial communications link.

When sufficient data have been collected, speech data collection and preprocessing can be stopped by a message from the voice processor. This is the normal way to stop data collection. Otherwise, the terminal processor will stop data collection after 400 samples have been collected (4 seconds). At the end of data collection, an end-of-speech message is sent to the voice processor.

3. COMMUNICATIONS

The communications portion of the terminal processor software is responsible for receiving instructions from the voice processor and

transmitting data to it. The communication software is capable of operating with an imperfect link - that is, error detection and recovery are performed. All messages have a two byte checksum and a sequence number. Errors within a message are detected with the checksum, and lost messages are detected when sequence numbers are missing. All messages are acknowledged by the receiver, and the message data are retained by the sender until the message has been acknowledged. Thus, if a message has an error or is lost, it can be retransmitted.

The communications processing in the terminal processor operates through several queues, which are first-in, first-out buffers implemented with linked lists. Each queue in the system has a specific function and contains messages that either have been received or that are being sent.

Data being sent to the voice processor are placed on a queue. When the data reaches the head of the queue, they are assigned a sequence number and transmitted. The block of data is then moved to another queue to await acknowledgment. If the data are not received correctly, they are retransmitted. When the data have been successfully received by the voice processor and acknowledged, the buffer containing the data is removed from the acknowledgment queue and made available for more data.

Data received from the voice processor are also queued. If the message is correctly received, it is queued until all lower sequence numbers are received. This assures that the data are processed in the same order that it is sent. When all lower numbered messages are accounted for, the message is placed on a queue based on its opcode and is removed by the task that processes that type of message.

A message begins with a header that controls routing and error detection. The header consists of four, one-byte fields. The first byte specifies the type or function of the message. A message may be data, an acknowledgment of a previously transmitted message, a retransmission of lost data, or a request for retransmission. Only data messages are acknowledged. The second byte contains the sequence number of the message, the value of which ranges from 0 to 127.

The third byte of the header contains the length of the message, including the header. The last byte of the header is the message opcode. This opcode specifies the function of the message and is used to determine which task within the terminal processor is to receive it. Each task in the terminal processor looks in a queue for incoming messages. The opcode byte in the message header is used to determine the queue in which to place it.

Each message ends with a two byte checksum. Each byte is an independent checksum using a different algorithm, providing extra error detection capability. When a message has been successfully received, an acknowledgment is immediately queued to be sent, and the message is placed on a queue to be processed.

During initial debugging of the system, ACKs from the voice processor to the terminal processor were being lost, due to high priority speech processing. The device service routine (DSR) that collects speech data requires more than 1 millisecond to collect data from all 14 filters. This caused the communications system to miss characters occasionally. It was decided to reassign the interrupt priority levels in the terminal processor to allow processing of a communications input character to interrupt the reading of the filter bank. This causes an occasional, approximately 0.3 ms skew between two parts of the spectrum, which is sampled once every 10 ms. This skew is insignificant, and the alteration in priority greatly improved communication throughput by reducing errors and hence retransmission of data.

Communications:	level 3 (highest)
Speech input:	level 4
Clock:	level 5
Portal/scale:	level 9
733 terminal:	level 12 (lowest)

Since the communications interrupt now has the highest priority, the communications driver in the terminal processor was partitioned to reduce the maximum elapsed time between the checking for other interrupts in order to minimize the delay caused by the interrupt processing.

An additional enhancement to the voice/terminal processor communications was the addition of an "inquire" type of message from the terminal to voice processor. Since the speech data blocks are quite long (88 bytes: header information, a two byte check sum, and 8 samples [80 ms] of speech data), spontaneous retries of *unacknowledged data* from the terminal processor significantly increased the traffic on the communication line. To avoid this, timeouts of unacknowledged messages cause the transmission of an "inquire" message of only six bytes. Receipt of this message at the voice processor will cause either an ACK or a NAK to be returned, with the terminal then taking the usual appropriate action.

4. PORTAL CONTROL

The portal control task is responsible for control of the booth. It monitors the door sensors, controls the locks and reads the weight scale. When there is a significant change in the state of the booth, a message is sent to the voice processor. For example, the detection of weight on the scale causes a message to be sent to the voice processor, which then requests that the "HELLO" message be spoken.

The portal control task is implemented as a finite state machine. The state transition tables are stored in the terminal processor. These tables determine the valid states and responses of the system. For example, the entry door is opened by a sequence that requires that the access door be locked before the entry door can be unlocked. This assures that the booth is always in a valid state and improves security.

5. LOCAL TERMINAL INTERFACE

The terminal processor contains a driver for a local terminal. This terminal is connected to an interface in the 990/5 and is used by the terminal processor to print statistics for performance measurement.

The keyboard on the local terminal is periodically examined by the terminal processor to accept commands to print debugging and performance information. The system can print lists of the currently running tasks and their states; communications statistics, including the number of retries; and speech statistics. Software switches are present that can be toggled through the keyboard to cause dynamic performance information to be printed. This performance information includes the average time to verify and the amount of data collected and processed for each phrase.

D. DOWNLOADING THE VOICE PROCESSOR AND THE TERMINAL PROCESSOR

The loaders are used to initially load the voice processor and terminal processor. All code for the system is stored on disk files on the host and is downloaded under operator control. Since a direct memory access channel exists between the host and the voice processor, this memory port is used to load the voice processor. The host reads the object code from disk, performs relocation and loads the binary object code directly into the voice processor's memory.

The command to load the voice processor is: LOADVP. The operator provides the name of the file on the host containing the object code for the voice processor. This code is then loaded into the voice processor's memory.

Since the only connection between the terminal processor and the host is through the voice processor, the terminal processor is loaded by the voice processor. The data to load are obtained from a file on the host, passed to the voice processor, and then passed on to the terminal processor.

When the LOADTP command is given by the operator, the host first loads the voice processor with a loader program for the terminal processor. It is then necessary for the operator to start this program executing via the front panel on the voice processor. The host provides the data across the TILINE* coupler to memory in the voice processor. The loader program in the voice processor adds checksum and header information and transmits the data to the terminal processor via the serial communications interface. The positive acknowledgment required from the terminal processor and the two bytes of checksum sent with each block of data assure that the information has been correctly received. When the data are acknowledged, the voice processor acknowledges it to the host by setting a flag in memory, and the host provides another block of data to the voice processor.

*Trademark of Texas Instruments Incorporated

The terminal processor has a loader present in read only memory. This loader was modified from the normal front panel loader as part of this project. The loader communicates through the serial communications interface on the terminal processor. Data is loaded in blocks with each block containing sequence number information and checksums. As each block is received, it is checked for validity. Data blocks are acknowleged by returning a message to the voice processor. This allows reliable loading of the terminal processor across an imperfect communications link. Loading of the terminal processor is initiated by pressing the halt and load buttons either on the front panel of the terminal processor or the buttons located remotely (at the host) and attached in parallel with the terminal processors front panel.

Once the terminal processor is loaded, the loader in the voice processor halts, and the host loads the voice processor's memory. The voice processor must then be started from its front panel, which completes the loading process.

SECTION IV

BISS VOICE AUTHENTICATION ALGORITHM MODIFICATIONS

In order to provide a baseline for discussion of the modifications made to the algorithm delivered on the BISS-ASV-ADM, the algorithm as defined in the "Segment Specification for Entry Control System" [11] has been reproduced as Appendix III of this final report. Rather than giving line-by-line modifications to Appendix III, narrative descriptions of the changes are given in this section. A brief summary of the differences among the BISS system, the 980-based CIC system, and the DSG and VVU systems is given in Tables 5 through 8.

A. WORD SET

The set of prompting words used for the old 980-based CIC system, for the laboratory test data set, for the DSG system, and for the VVU system is shown below. It is different from that given in Appendix III for the BISS system.

GOOD	BEN	SWAM	NEAR
PROUD	BRUCE	CALLED	HARD
STRONG	JEAN	SERVED	HIGH
YOUNG	JOYCE	CAME	NORTH

While the words are still selected randomly during verification, the prompting order during enrollment has been fixed for both the DSG and VVU systems. Since the need for randomization (to prevent the use of tape recorders by impostors) is not applicable, fixing the phrase order during enrollment allows the enrollee to use a printed list for assistance in understanding the phrases, which not only are new to the user, but which have degraded in quality from the old BISS system due to the need to use LPC coded prompts, as explained earlier. The prompting order for the phrases during enrollment is given in Table 9.

Table 5

ASV System Comparisons

Overall

	BISS-ADM	CIC	DSG (VVU)
Computer:	980B	980A	990/10,/10 (990/10,/12,15)
Operating System:	Modification of RTM-I	RTM-I	DX-10, TX-990 (DX-10, TX-990, TIPMX)
Prog. Languages:	Fortran, 980 Assem	Fortran, 980 Assem	PASCAL, 990 Assem
Entry Control:	No Booth Checks: ● None	Booth with Scale ● Multiple Entrants Checks: ● Security Level	Booth with Scale ● Multiple Entrants Checks: ● Security Level ● Time of Day ● Inventory
Training Mode (Mini-Enrollment)	No	No	No
Validation Mode (Verify; No Statistics)	No	No	No
Statistics/Report Generation	None	Some	Extensive

Table 6

ASV System Comparisons

Speech Input

	BISS-ADM	CIC	DSG (VVU)
Prompting Set:	"North Lawn Great Camp"	"Good Bruce Called Hard"	"Good Bruce Called Hard"
Filters:	Digital	Digital	Analog
	• Prog. Gain Control	• No Prog. Gain Control	• No Gain Control
	• Overload Flag	• No Overload Flag	• No Overload Flag
	• Constant CF Spacing	• Constant CF Spacing	• CF Spacings Increase with Frequency
	• Constant BWs	• Constant BWs	• BWs Increase with Frequency
	• 16 Filts (Discard 15,16)	• 16 Filts (Discard 15,16)	• 14 Filters
Preprocessing:			
• Regression Vectors:	Legendre Polys.	Both BISS and	SIN/COS
• Regression Limiting:	No	DSG-Type Preprocessings	Yes
• Energy:	Max of 14 Filters	Used; Energy Calc.	Max of Filts (2-13)-Ave Filts (1-14)
• Normalization:	By the Mean	only DSG Type	by the St. Dev.
• Quantization Thresholds:	All Male Design Set Discard Lowest 4 Filters		Male/Female Design Set Filters Selected by Vowel

Table 7

ASV System Comparisons

Enrollment

	BISS-ADM	CIC	DSG (VVU)
Phrase Prompting Order:	Random	Random	Fixed
Auto Termination:	No; Operator Controlled	Yes; Variety of Criteria	Yes; Variety of Criteria
Max Acceptable e:	∞	160 (BISS) 200 (DSG)	200
Reference File Precision per Element (Frac. Bits for Updating)	3 Int. 5 Frac.	3 Int. 5 Frac.	3 Int. 4 Frac.

Table 8

ASV System Comparisons

Verification

	BISS-ADM	CIC	DSG (VVU)
Random No. Generation for Phrase Select:	Mult. Congruential	Linear Congruential	Linear Congruential
Sequential Strategy:	$\leq$ 4 Phrases • All Different • Auto Retry	$\leq$ 7 Phrases • 4 Different • No Auto-Retry	$\leq$ 7 Phrases • 4 Different • No Auto-Retry
Unregistered Phrases:	Reprompt Immediately	Reprompt after 4	Reprompt after 4
Decision FCN:	$D = \dfrac{\sum_{k=1}^{N}\sum_{i=1}^{4} e_{ik}}{\mathrm{Max}[\mathrm{Min}(4N\hat{e}_{max}, \sum_{k=1}^{N}\sum_{i=1}^{4}\hat{e}_{ik}), 4N\hat{e}_{min}]}$	$D = \dfrac{\sum_{k=1}^{N}\sum_{i=1}^{4} e_{ik}}{\sum_{K=1}^{N}\sum_{i=1}^{4}\hat{e}_{ik}}$	$\dfrac{\sum_{k=1}^{4}\sum_{i=1}^{4}\hat{e}_{ik}}{\max[\min(16\hat{e}_{max}, \sum_{k=1}^{4}\sum_{i=1}^{4}\hat{e}_{ik}), 16\hat{e}_{min}]}$
$\hat{e}_{max}$, $\hat{e}_{min}$:	Fixed	Function of Session & Phrase Number	Function of Session & Phrase Number
Decision Thresholds:	Function of Session & Phrase Number & Mode	Function of Session & Phrase Number & Mode	Function of Session & Phrase Number & Mode
Updating	1/4; 1/16 (1-4); (> 4)	1/4; 1/8; 1/16 (1-4); (5-12); (>12)	1/4; 1/8; 1/16 (1-4); (5-12); (>12)

TABLE 9. PHRASE PROMPTING ORDER DURING ENROLLMENT

1.	PROUD BEN SWAM NEAR	17.	STRONG JOYCE CAME NEAR
2.	GOOD BRUCE CALLED HARD	18.	GOOD BRUCE CALLED HIGH
3.	STRONG JOYCE CAME NORTH	19.	YOUNG JEAN SERVED HARD
4.	YOUNG JEAN SERVED HIGH	20.	PROUD BEN SWAM NORTH
5.	STRONG JOYCE CALLED HARD	21.	GOOD BRUCE CAME NEAR
6.	GOOD BRUCE CAME NORTH	22.	PROUD BEN SERVED HARD
7.	YOUNG JEAN SWAM NEAR	23.	YOUNG JEAN SWAM NORTH
8.	PROUD BEN SERVED HIGH	24.	STRONG JOYCE CALLED HIGH
9.	YOUNG JOYCE CAME NORTH	25.	GOOD BEN SWAM NORTH
10.	GOOD BEN SWAM NEAR	26.	YOUNG JOYCE CAME NEAR
11.	STRONG JEAN SERVED HIGH	27.	STRONG JEAN SERVED HARD
12.	PROUD BRUCE CALLED HARD	28.	PROUD BRUCE CALLED HIGH
13.	STRONG JEAN SWAM NEAR	29.	STRONG JEAN SWAM NORTH
14.	YOUNG JOYCE CALLED HARD	30.	GOOD BEN SERVED HARD
15.	PROUD BRUCE CAME NORTH	31.	PROUD BRUCE CAME NEAR
16.	GOOD BEN SERVED HIGH	32.	YOUNG JOYCE CALLED HIGH

B. SPEECH PREPROCESSING

Except for the LPC-based experiments described in Section VII, all the speech processing algoritms described in this report are based on the relative spectrum of speech as a function of time, as represented by the energies out of a bank of bandpass filters, sampled every 10 ms, which has been preprocessed as described in this subsection. In order to simplify the notation, the time index (j) has been eliminated from the equations in this section.

1. FILTER BANK DEFINITIONS

This section will provide a superficial distinction between some of the filter banks used in recent speaker verification work at Texas Instruments.

The first column in Table 10 contains the design center frequencies (165.3 Hz spacing) for the two-stage (two poles per stage) digital filters used in each of the two filter banks delivered with the BISS ADM. For the BISS ADM the center frequencies of each of the two stages were slightly offset to produce a rather flat-topped amplitude response. The sample period for the actual filters delivered, however, was 106 instead of 100 microseconds, resulting in center frequencies about 6% lower than those in the design. These lower center frequencies were actually used in all of the MITRE testing and still reside in the ADM.

For the Total Voice contract, the center frequencies and bandwidths of the filters in one of the filter banks (Channel #2) in the BISS ADM were changed by replacing the PROMs which contained the filter coefficients. The Total Voice filter coefficients were defined such that the center frequencies of each of the two stages were the same for each of

the filters. This yielded round-shouldered filter amplitude responses. The same 106 vs 100 microsecond sample period disparity remained, however, yielding the measured values shown in Table 10.

The DSG filter banks (also used on the VVU system) were an outgrowth of the Total Voice filters. The first step was to define a filter bank having a set of iteratively defined center frequencies and bandwidths. After this was done, experiments to be described in the next section showed that speaker verification performance improved as the bandwidths of the filters decreased. The result was the 14-channel DSG filter bank, as defined by the first fourteen entries in the final column of Table 10. The iterative definition of these parameters is given by

$$BW_{n+1} = BW_n \, (2)^{1/13} \qquad CF_{n+1} = CF_n + (CF_n - CF_{n-1}) \, (2)^{1/6}$$

where, $BW_1 = 200$, $CF_1 = 350$, and $CF_0 = 350 - 100/(2)^{1/6}$

TABLE 10. FILTER BANK DEFINITIONS

	BISS ADM DESIGNED		TOTAL VOICE DESIGNED		TOTAL VOICE MEASURED		DSG/VVU	
FILTER #	CF	BW	CF	BW	CF	BW	CF	BW
1	410	290	350	300	280	250	350	200
2	575	290	450	300	395	280	450	211
3	741	290	555	310	525	310	562	223
4	906	290	670	340	630	340	688	235
5	1071	290	790	380	750	360	830	248
6	1237	290	940	400	900	360	988	261
7	1402	290	1120	400	1080	360	1167	275
8	1567	290	1320	400	1265	365	1367	290
9	1733	306	1550	400	1480	365	1591	306
10	1898	290	1810	400	1725	365	1843	323
11	2063	290	2100	400	1985	365	2126	341
12	2229	290	2430	400	2285	360	2443	360
13	2394	290	2800	400	2640	365	2800	379
14	2559	290	3350	700	3150	625	3200	400
15	*2725	290	4000	700	3720	635	**3649	422
16	*2890	290	4650	700	4235	615	***4153	445

* Not used in ADM, although implemented in actual filters

** As defined for CCD analyzer; does not exist in DSG filter board.

*** Cut-off freq for high pass filter as defined for CCD analyzer; does not exist on DSG board.

2. REGRESSION

It has been found that by eliminating the gross aspects of the spectrum, such as the slope and curvature, more clearly defined formant frequencies are obtained. This can be accomplished by regression. The BISS-ADM utilized the first two Legendre polynomials for regression; however, the DSG and VVU systems regressed the spectral amplitude vector using the first three elements of an orthonormal sin/cosine basis set, defined as follows:

$$r_0(i) = \frac{1}{\sqrt{14}} \qquad r_2(i) = \frac{1}{\sqrt{7}} \text{ cosine} \left[\frac{(i-0.5)}{14}\pi\right]$$

$$r_1(i) = 0.559065 - 0.8763356 \sin\left[\frac{(i-0.5)}{14}\pi\right] .$$

The regression coefficients are then defined as

$$C_k = \sum_{i=1}^{14} a(i) * r_k(i) \qquad k = \{0,1,2\}$$

where a(i) is the magnitude of the output of filter (i). [The BISS specification in Appendix III denotes this as f(i).] The regressed output (denoted g(i) in Appendix III) is then determined from

$$a_R(i) = a(i) - [C_0 * r_0(i) + C_1 * r_1(i) + C_2 * r_2(i)] .$$

In the actual implementation, however, scaling such as described in section 10.2.2.2.1.1 of Appendix III is done to preserve precision in the calculation. This is accomplished by redefining C_1 and C_2 as

$$C'_1 = \sum_{i=1}^{14} \frac{a(i) * phi_1(i)}{32768} \qquad C'_2 = \sum_{i=2}^{14} \frac{a(i) * phi_2(i)}{32768}$$

and

$$a_R(i) = a(i) - C_0 * r_0(i) - \frac{C'_1 * gamma_1(i)}{32768} - \frac{C'_2 * gamma_2(i)}{32768},$$

where

$$phi_1(i) = \frac{32768 * 32768}{0.5 * 30211} * r_1(i) \qquad phi_2(i) = \frac{32768 * 32768}{0.5 * 24615} * r_2(i)$$

$$gamma_1(i) = 0.5 * 30211 * r_1(i) \qquad gamma_2(i) = 0.5 * 24615 * r_2(i)$$

The phis and gammas are given in Table 11, which corresponds functionally to Table VII in Appendix III.

TABLE 11. REGRESSION VECTORS

i	$phi_1(i)$	$phi_2(i)$	$gamma_1(i)$	$gamma_2(i)$
1	32767	-32767	6963	-4622
2	19168	-31125	4073	-4390
3	6603	-27920	1403	-3938
4	-4304	-23316	-914	-3288
5	-13000	-17543	-2762	-2474
6	-19056	-10891	-4049	-1536
7	-22158	-3692	-4708	-520
8	-22158	3692	-4708	521
9	-19056	10892	-4049	1537
10	-13000	17543	-2762	2475
11	-4304	23316	-914	3289
12	6603	27920	1403	3939
13	19168	31125	4073	4391
14	32767	32767	6963	4623

Thus the regression tends to flatten the spectrum, removing any half-cycle sine or cosine wave trends of the spectrum. An example of a spectral waveform having a large positive C_1 is a nasal, which has one peak near the low end and one near the high end of the spectrum (around 250 Hz and 2200 Hz). An example of a spectral waveform with a large positive C_2 is a sibilant, having most of its energy above 3000 Hz. Most vowels, however, have the opposite spectral tilt because of the glottal source spectral decay with increasing frequency, yielding a large negative value of C_2.

It has been observed, however, that regression sometimes eliminates too much of the variance of the spectrum. Hence, it was decided that if the postregression variance were less than some fraction of the preregression variance, the magnitudes of the regression coefficients would be reduced sufficiently to limit the variance after regression to that fraction of the preregression variance. Stated another way, the RMS value of the amount regressed out is limited to some (other) fraction of the preregression variance. If in fact the regression coefficients are too large, they are reduced as follows:

$$C_{1R} = C_1 * \frac{K * (\text{pre-regression variance})}{\text{sqrt}(C_1^2 + C_2^2)}$$

where

$$\text{pre-regression variance} = \text{sqrt}\left(\sum_{i=1}^{14} a(i) - C_0^2 \right)$$

$$K = \text{sqrt}\left(1 - \frac{13}{11} * \text{ratio}_{min}^2 \right)$$

and ratio_{min} is the minimum allowable ratio of postregression variance to preregression variance. In actually implementing the regression limiting, intermediate scalings were performed using C_1' and C_2' to preserve as much precision as possible. Although the need to limit regression may exist for regression using any set of basis vectors, such limiting was done only when sine/cosine basis functions were used. Regression limiting was not done for the original BISS algorithm, which used Legendre polynomial basis functions.

3. NORMALIZATION

The next step in the preprocessing is to normalize the regressed values of the filter outputs to accommodate the wide variety of vocal efforts among speakers. In the original BISS algorithm, normalization of each spectral section was done by dividing by the mean of all the filter outputs in the spectral section. However, when the regression was done using the sine/cosine basis functions, normalization was done by dividing by the postregression standard deviation, given by

$$\text{postregression variance} = \frac{1}{11} \left(\sum_{i=1}^{14} a(i) - C_0 - C_{1R} - C_{2R} \right)$$

plus a constant that was added to avoid large normalized amplitudes during times of very small standard deviations (i.e., during silence intervals). Whenever regression limiting occurred, the postregression standard deviation became the ratio ${}_{min}$ (discussed in the prior subsection) times the preregression standard deviation.

4. QUANTIZATION

The regressed, normalized filter (i) energies for a spectral section are next quantized to one of eight levels according to a set of quantization thresholds, $\phi_{i,q}$:

$$
a(i)_Q = \begin{cases} 0 \text{ for } a(i)_N \leq \phi_{i,0} \\ q \text{ for } \phi_{i,q} < a(i)_N \leq \phi_{i,q+1} & \text{for } q = \{1,2,\ldots 6\} \\ 7 \text{ for } \phi_{i,7} < a(i)_N & . \end{cases}
$$

The selection of the thresholds used is clearly a function of the processing that precedes this quantization step, i.e. the center frequencies of the filter bank, the regression, and the normalization. Hence, for each change in any of these steps, the quantization levels must be recalculated.

In originally determining the quantization thresholds used for BISS, a special data set was used, consisting of one repetition by each of 21 males of the set of words given in Table 12.

TABLE 12. WORD SET #1 USED IN DETERMINING QUANTIZATION THRESHOLDS

Bid	Bed	Bored
Bead	Bade	Bode
Bide	Booed	Bud

However, since this set of words did not include all of the vowels, and since there were no females in the data set (a possible reason for poorer performance for females in prior speaker verification experiments), a new set was collected consisting of one repetition by each of 11 males and 11 females of the set of words given in Table 13.

TABLE 13. WORD SET #2 USED IN DETERMINING QUANTIZATION THRESHOLDS

*Pot	*Bert	*Bet	*Bought
*Put	Bout	Bait	*Beet
Boyd	*Bat	Boat	*But
Butte	Bite	*Bit	*Boot

The quantization thresholds were then chosen by plotting histograms for each of the regressed, normalized filter outputs $[a(i)\text{'s}]_N$ during the vowel portions of the words. Rather than having these quantization levels chosen to yield a uniform probability, it seemed more desirable to cluster the quantization thresholds at higher energy levels. In this way the sensitivity to noise can be reduced and the quantization resolution increased. Two schemes were used to exclude the low-energy filters. One was simply to exclude the four lowest energy filters at each time sample during the vowel. (This was the scheme used to determine thresholds for the original BISS algorithm.) Another method was to more selectively include only energies from certain filters on the basis of the vowel's identity. In this second method, since the identities of the high energy filters are a function of the center frequencies of the particular filter bank used, the high amplitude filters were determined from visual inspection of spectrograms. The resulting choices for high amplitude filters are given in Tables 14 and 15 for the BISS and for the VVU filter banks, respectively.

In an effort to compare both the two methods of generating quantization thresholds and the two data sets, the words with common vowels from both data sets were processed using the BISS filter simulation and IPMOD2 preprocessing. Figures 18A and 18B compare the old and the new data sets using only the male speakers and the ten highest amplitude filters for each word. Figures 18A and 18B compare the resulting thresholds for the new data set using the ten highest amplitude filters for males alone (18A) to those for both males and females (18B). Figures 18B and 18D compare (for all speakers in the new data set) using the ten highest amplitude filters to using selected high amplitude filters, as given in Table 14.

In all subsequent experiments discussed in Sections V and VI, the new data set, with both male and female speakers, was used to determine quantization thresholds using the selected filters of Tables 14 and 15. However, in these subsequent experiments, only the pure vowels (* in Table 13) were used.

The particular quantization thresholds used in the VVU system are given in Table 16, which corresponds functionally to those given in the BISS specification (Appendix III). The differences in the conditions under which these two sets of thresholds were determined are shown in Table 17.

Table 14

High Amplitude Filters for BISS Filter Bank

Vowel	Word		Filter Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14
i	beet	Male	X	X							X	X	X	X	X	X
		Female	X	X										X	X	X
ɪ	bit	Male		X	X					X	X	X	X	X	X	X
		Female		X	X						X	X	X	X	X	X
ɛ	bet	Male		X	X					X	X	X			X	X
		Female		X	X	X					X	X	X	X	X	X
æ	bat	Male		X	X	X				X	X	X			X	X
		Female		X	X	X	X				X	X	X	X	X	X
ɑ	pot	Male			X	X	X	X	X	X					X	X
		Female			X	X	X	X			X	X				X
ɔ	bought	Male		X	X	X	X	X							X	X
		Female		X	X	X	X	X								X
ʌ	but	Male		X	X	X			X	X					X	X
		Female		X	X	X	X			X	X	X	X	X		
ʊ	put	Male	X	X				X	X	X					X	X
		Female		X	X					X	X	X				X
u	boot	Male	X	X				X	X	X			X	X	X	X
		Female		X						X	X	X			X	X
ɝ	Bert	Male	X	X				X	X	X						
		Female		X	X				X	X	X	X				

Table 15

High Amplitude Filters for TVP/VVU Filter Bank

Vowel Word			Filter Number													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
i	beet	(M)*	X	X								X	X	X	X	X
		(F)	X	X	X									X	X	X
ɪ	bit	(M)		X	X	X					X	X	X	X	X	
		(F)		X	X	X						X	X	X	X	X
ɛ	bet	(M)		X	X	X					X	X		X	X	
		(F)		X	X	X	X					X	X	X	X	X
æ	bat	(M)			X	X	X				X	X		X	X	
		(F)			X	X	X	X				X	X	X	X	X
a	pot	(M)				X	X	X	X	X	X			X	X	X
		(F)				X	X	X	X			X			X	X
ɔ	bought	(M)			X	X	X	X	X					X	X	
		(F)			X	X	X	X	X						X	
ʌ	but	(M)			X	X	X			X	X			X		X
		(F)			X	X	X	X			X	X			X	
U	put	(M)	X	X	X				X	X	X			X		
		(F)			X	X					X	X			X	
u	boot	(M)	X	X					X	X	X		X	X		
		(F)		X	X						X	X		X	X	
ɝ	Bert	(M)	X	X	X				X	X	X					
		(F)		X	X	X				X	X	X				

*M: Male Speakers, F: Female Speakers

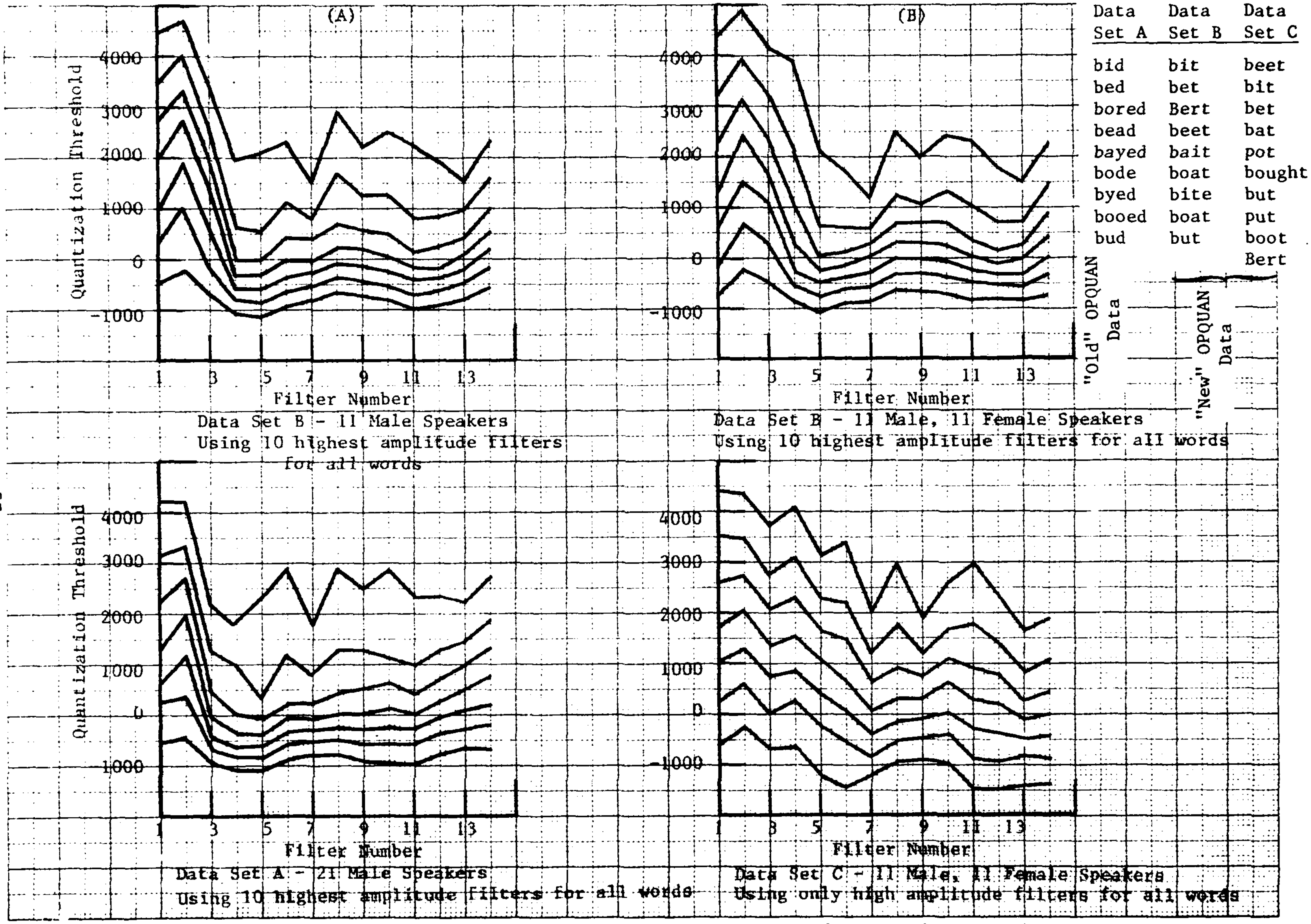


Figure 18 Quantization Threshold Comparisons Using BISS Filter Simulation and IPMOD2 Preprocessing

<u>**Table 16**</u>

QUANTIZATION THRESHOLDS
DSC FILTER BANK
NORMALIZING BY STANDARD DEVIATION

Filter Number	Quantization Level						
	1	2	3	4	5	6	7
1	-2330	-1489	-761	10	583	1120	1940
2	-76	531	1222	1844	2529	3195	4088
3	-390	59	577	1103	1744	2518	3424
4	-206	479	1071	1704	2353	3062	4047
5	-194	420	1010	1586	2086	2854	3851
6	-998	-139	496	1032	1629	2246	2990
7	-1466	-561	-30	501	1006	1753	2660
8	-1020	-610	-165	278	795	1518	2522
9	-1258	-750	-283	136	681	1539	2437
10	-1282	-749	-164	364	1139	1931	2863
11	-568	174	796	1396	2010	2702	3750
12	-374	188	744	1312	1934	2813	3760
13	-660	-167	260	726	1231	1871	2749
14	-1933	-1298	-830	-453	-57	468	1113

TABLE 17. CONDITIONS FOR DETERMINING QUANTIZATION THRESHOLDS

Condition	BISS	VVU
Filter bank (Table 10)	BISS filters	DSG filters
Data set as per	Table 12	Table 13 (only 10 vowels)
Speaker population	21 males	11 males, 11 females
Filters used	10 high energy	Selected as per Table 15
Regression	Legendre polyn.	Sine/cosine
Regression limiting	no	yes
Normalization by	mean	standard deviation

5. ENERGY

For each time sample, a measure of the energy was also computed. In the BISS algorithm, this was just the maximum energy from the fourteen filters. As an aid to distinguishing vowels from nasals [which usually have most of their energy in a(1)] and vowels from sibilants [which usually have most of their energy in a(14)], these two filters were not used in calculating the new energy measure given by

$$E_1 = \mathrm{sqrt}\left[\sum_{i=2}^{13} a(i)^2 - \frac{1}{11} \left(\sum_{i=2}^{13} a(i) \right)^2 \right] \qquad \text{(Def. 1)}$$

During the experiments described in Section V, it was determined that the following energy definition gave still better definition of the vowel boundaries.

$$E_2 = \frac{1}{8} \left[14 * \max_{i=2,13} \{ a(i) \} - \sum_{j=1}^{14} a(j) \right] \qquad \text{(Def. 2)}$$

C. VERIFICATION PROCESSING

After the input speech data have been preprocessed, they are compared to reference patterns for the four given words by "scanning" the reference patterns across the input data and calculating a "scanning error" for each centisecond of data. Minima (valleys) in these scanning errors are saved and an optimal sequence of valley points is found using a partial sequence error (called a "point-pair error") between adjacent valleys. After an optimal sequence of valley points is found, a decision function is calculated and is compared to a decision threshold. This subsection will describe the changes that have been made in the point-pair error calculation, in the decision strategy, in the decision function calculation, and in various parameters used in these calculations.

1. POINT-PAIR ERROR CALCULATION

This subsection gives the equations used in the point pair error calculations for various systems.

<u>BISS:</u>

$$EW = \frac{(e(i) + 1) * (e(i+1) + 1)}{2048} \left(1 + \frac{|dt - dthat|}{\max(dthat, \min dthat)}\right)$$

where: .5 x dthat < dt < 2 x dthat; min dthat = 20; max EW =70;

<u>CIC:</u>

$$EW = \left(\frac{e(i)}{10} + 1\right)\left(\frac{e(i+1)}{10} + 1\right)\left(1 + \frac{|dt - dthat|}{\max(dthat, \min dthat)}\right)$$

where: .5 x dthat < dt < 2 x dthat; min dthat = 20; max EW = 1500

<u>DSG/VVU:</u>

$$EW = \frac{(e(i) + K) * (e(i+1) + K)}{1024} \left(1 + \frac{|dt - dthat|}{\max(dthat, \min dthat)}\right)$$

where: .33 x dthat < dt < 2 x dthat; min dthat = 10; max EW = 200;
K = 100;

To check that the max EWs are consistent among all the equations, find the e(i)'s under the conditions of dt = dthat, e(i) = e(i+1), and EW = max EW. This yields e(i) = 377.3 for CIC, e(i) = 378.6 for BISS, and e(i) = 352.5 for DSG/VVU. This lower e(i) for the BISS/CIC system is due to the higher "K," which tends to decrease the sensitivity of EW to low valley point errors. The benefit of this decrease in sensitivity is to keep very low e(i)'s from allowing very high e(i+1)'s.

For reference purposes, the equations used in the two "Total Voice" programs are given below. These programs use the optimal sequence finder subroutine to find sequences of either 2 or 3 reference points in words, rather than to find sequences of 4 points in a phrase, as in the original BISS and VVU systems. The real difference, however, is that the format and size of the "scanning patterns" are different. The BISS system, the old CIC system, and both the DSG and VVU systems, use 6 time slices (samples) of 14 filters (84 elements total). The original total voice program used 2 columns of 9 filters, 2 regression coefficients and energy, plus the difference between these two 12 element columns (36 elements total). The revised total voice program used 5 columns of 14 filters, 2 regression coefficients and energy, plus the difference between adjacent columns (153 elements total). Hence, scanning errors will have a different range of values, as reflected by the chosen valley point error maxima: 200 for the original total voice, 615 for the revised total voice, and 400 (now 500) for BISS/CIC.

ORIGINAL TOTAL VOICE:

$$EW = \frac{(e(i) + K) * (e(i+1) + K)}{2048} \left(1 + \frac{|dt - dthat|}{max(dthat, min\ dthat)}\right)$$

where: min dthat = 0; K = 40; max EW, dtmin & dtmax are word dependent.

REVISED TOTAL VOICE:

$$EW = \frac{(e(i) + K) * (e(i+1) + K)}{1024} \left(1 + 2 * \frac{(dt - dthat)}{max(dthat, min\ dthat)}\right)^2$$

where: min dthat = 4; K = 100; max EW, dtmin & dtmax are word dependent

All of these equations can be rewritten in the following form, with the constants specified in Table 18. Note that if K is chosen correctly for a given filter bank and preprocessing, the C*EWMAX product remains constant.

$$EW = \frac{1}{C}\left(\frac{e(i)}{K} + 1\right)\left(\frac{e(i+1)}{K} + 1\right)\left(1 + \beta\ \frac{|dt - dthat|\alpha}{max(dthat, min\ dthat)}\right) .$$

TABLE 18. POINT-PAIR ERROR PARAMETERS

SYSTEM	C	K	β	α	min dthat	max EW
CIC	1	10	1	1	20	1500
BISS	2048	1	1	1	20	70
DSG	0.1024	100	1	1	10	70
OTV	1.28	40	1	1	0	*
RTV	0.1024	100	2	2	4	*

* Word dependent

One interpretation of K is for it to be the ehat averaged over the population (<ehat>). In trying to establish the values of the ehats averaged over the population, we can refer to the average ehats gathered from the CIC systems and from laboratory experiments, as given in Table 19.

TABLE 19. AVERAGE EXPECTED SCANNING ERRORS

		BISS FILTER BISS PREPRO	BISS FILTER IPM2 PREPRO	DSG FILTER IPM2 PREPRO
TEST SET	MALES	109	130	117
AFTER ENRLMNT	FEMALES	121	148	135
TEST SET	MALES	91	109	101
AFTER 21 SESS.	FEMALES	105	126	118
CIC OVERALL (MALES & FEMS)		94	125	94
VVU OVERALL (MALES & FEMS)		-	-	101

Since the average ehats clearly depend upon filter bank and preprocessing method, the Ks should also reflect this difference. Note that the average ehats are further dependent upon sex and word (reference Section IX), and hence consideration should be given in the future to at least making K dependent upon the user's sex.

By the interpretation of K as being <ehat>, the K of 100 used for the DSG/VVU system is certainly reasonable.

2. DECISION STRATEGY

In the BISS algorithm, one session of up to four phrases was used to make a decision to accept the user. If the person were not accepted, a second (independent) session of up to four phrases was used, with tighter thresholds than those used in the first session. However, the algorithm has since been revised so that rather than totally disregarding the first four phrases spoken by the user, either all spoken phrases (up to four) would be used in the decision process with normal decision thresholds or, alternately, only the "best" (of up to seven) phrases would be used, in which case, more restrictive (prejudiced) thresholds would be used in the decision.

Figure 19 is a flow chart of the revised algorithm. In the revised algorithm, after the Nth phrase is processed ($N < 4$), if all phrases have been "registered" (a match has been made between the input and the reference patterns for all of the prompted words), a decision function (see next subsection) is calculated using appropriate parameters and is compared to a "normal" mode decision threshold. If the decision function is less than the threshold, the user has been verified; if not, the worst phrase is excluded and the decision function is recalculated and compared to a "prejudiced" mode decision threshold. This procedure is repeated until either the user has been verified or all phrases have been excluded. Then, if the user has not been verified, if less than seven phrases have been prompted, and if no more than three phrases have been misregistered, another phrase is prompted and the decision procedure is repeated.

If N is larger than four, or if one or more of the phrases was not registered, the "prejudiced" mode thresholds are used from the start,

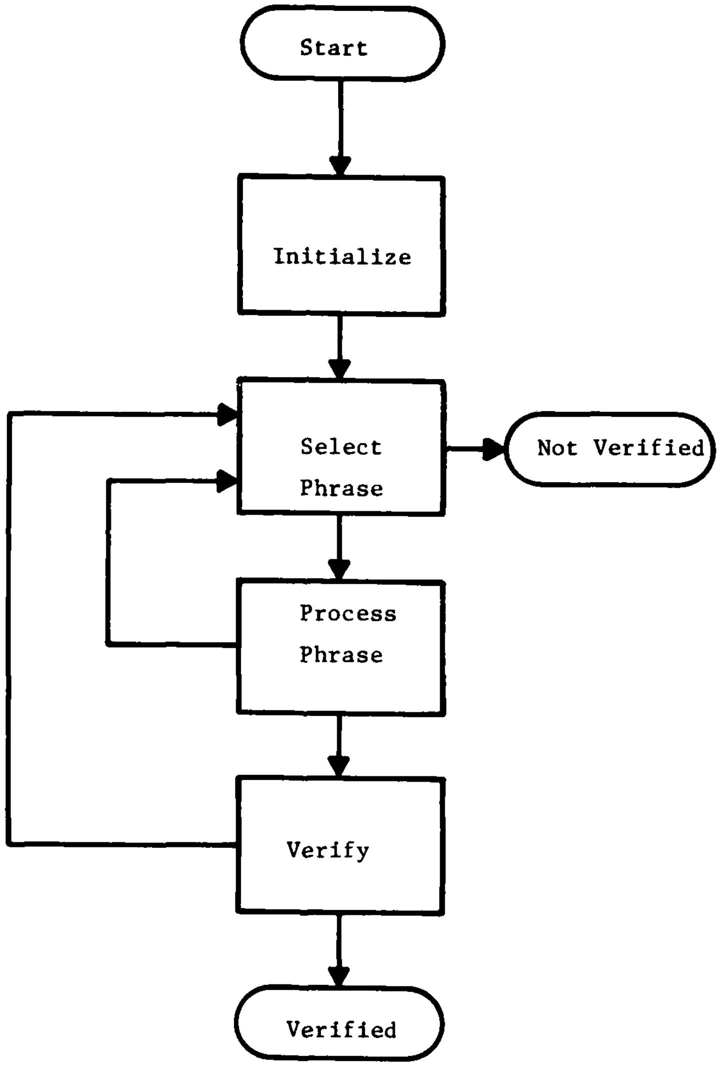
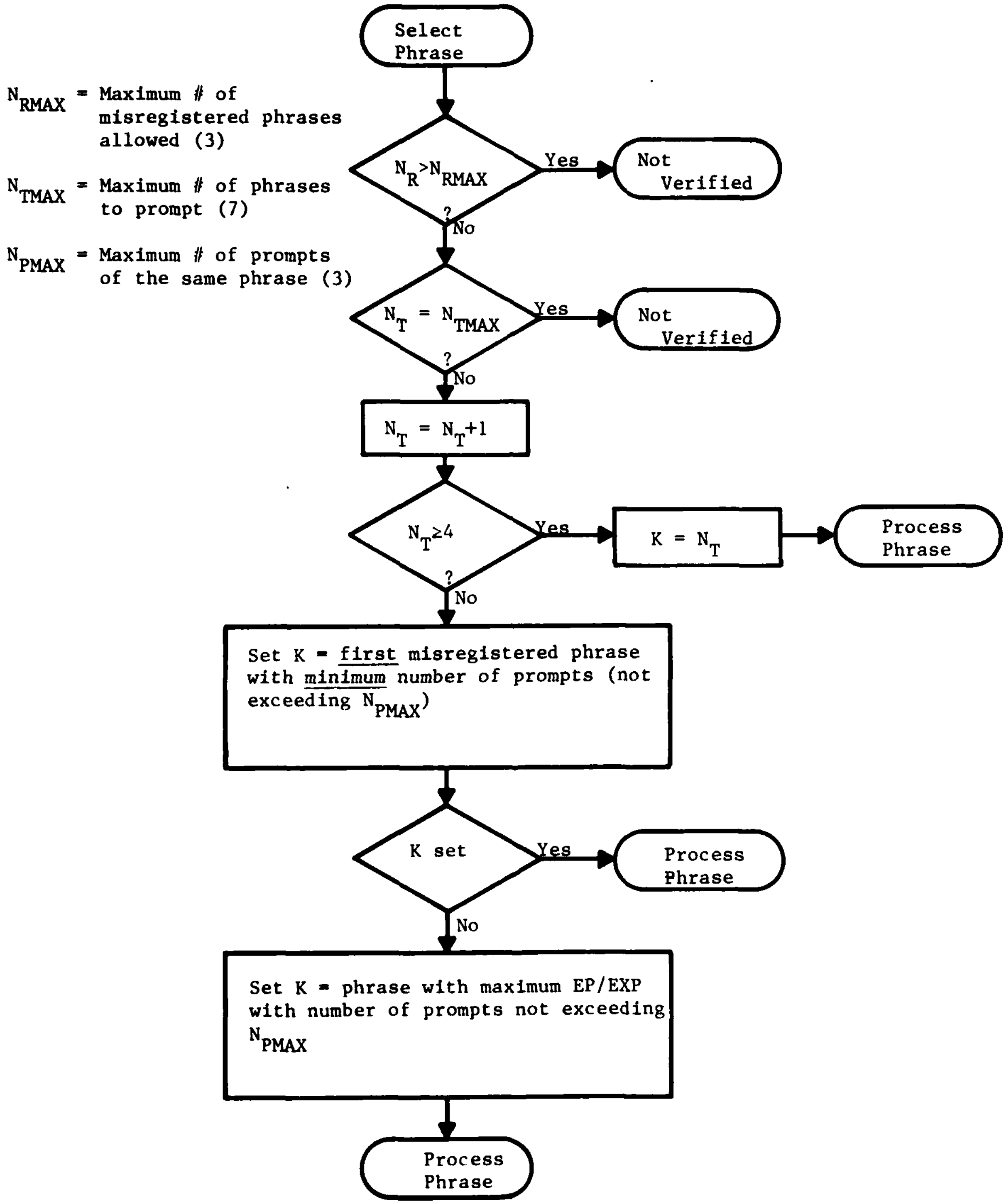
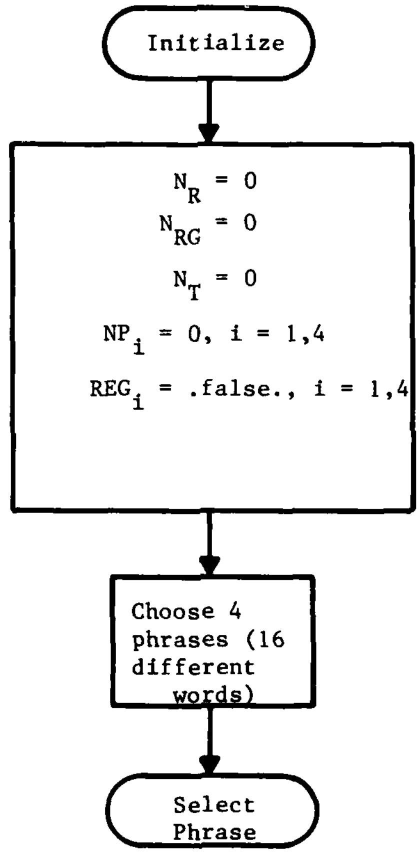


Figure 19 CIC Verification Strategy Overview


Select Phrase
N_RMAX = Maximum # of misregistered phrases allowed (3)
N_TMAX = Maximum # of phrases to prompt (7)
N_PMAX = Maximum # of prompts of the same phrase (3)
N_R>N_RMAX ?
Yes
No
Not Verified
N_T = N_TMAX ?
Yes
No
Not Verified
N_T = N_T+1
N_T≥4 ?
Yes
No
K = N_T
Process Phrase
Set K = first misregistered phrase with minimum number of prompts (not exceeding N_PMAX)
K set
Yes
No
Process Phrase
Set K = phrase with maximum EP/EXP with number of prompts not exceeding N_PMAX
Process Phrase



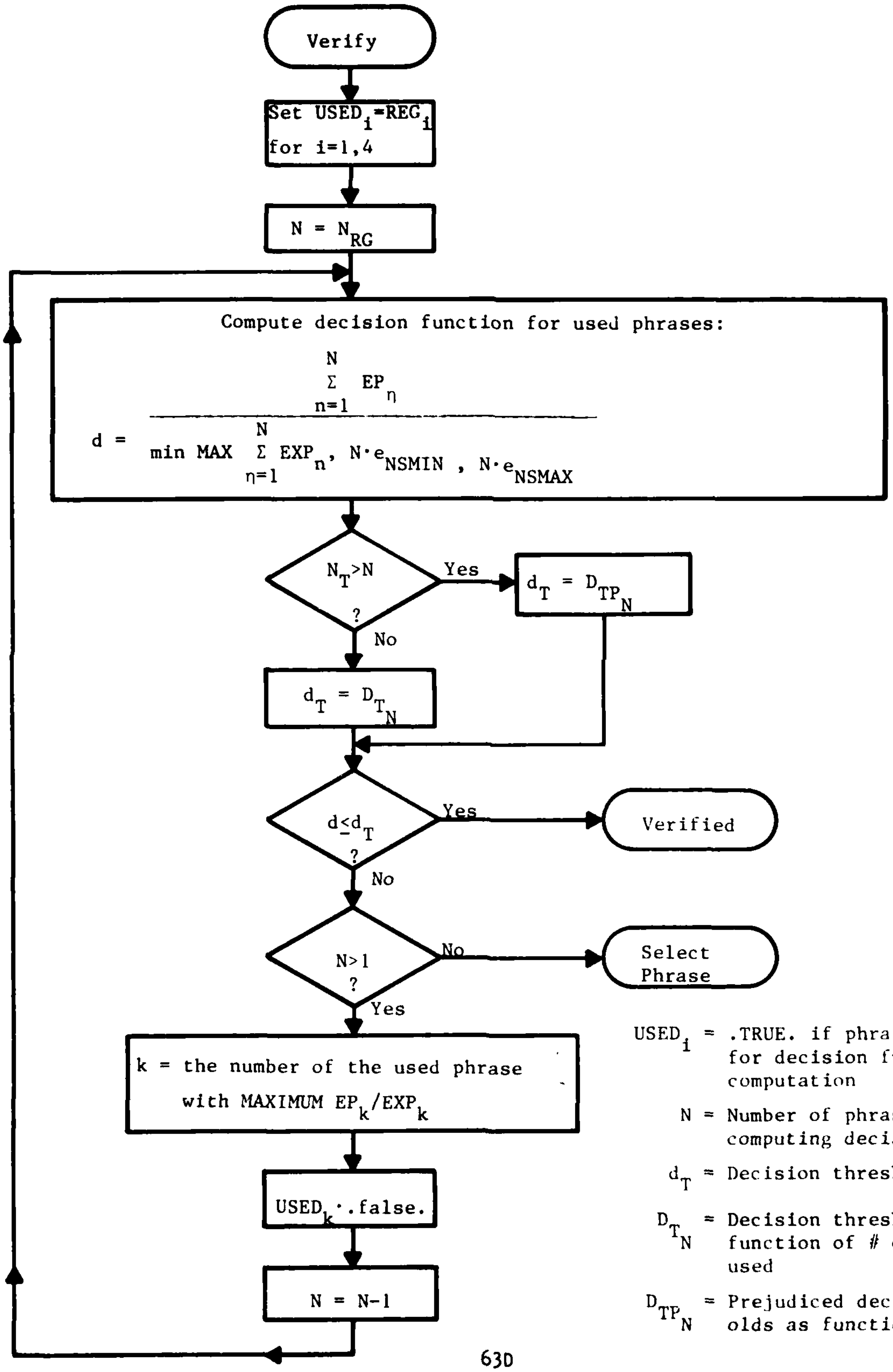
N_R = Number of phrases not registered

N_{RG} = Number of unique registered phrases

N_T = Total number of phrases prompted

NP_i = Total number of times of times the i^{th} phrase is prompted

REG_i = .True. when phrase i is registered, .false. otherwise



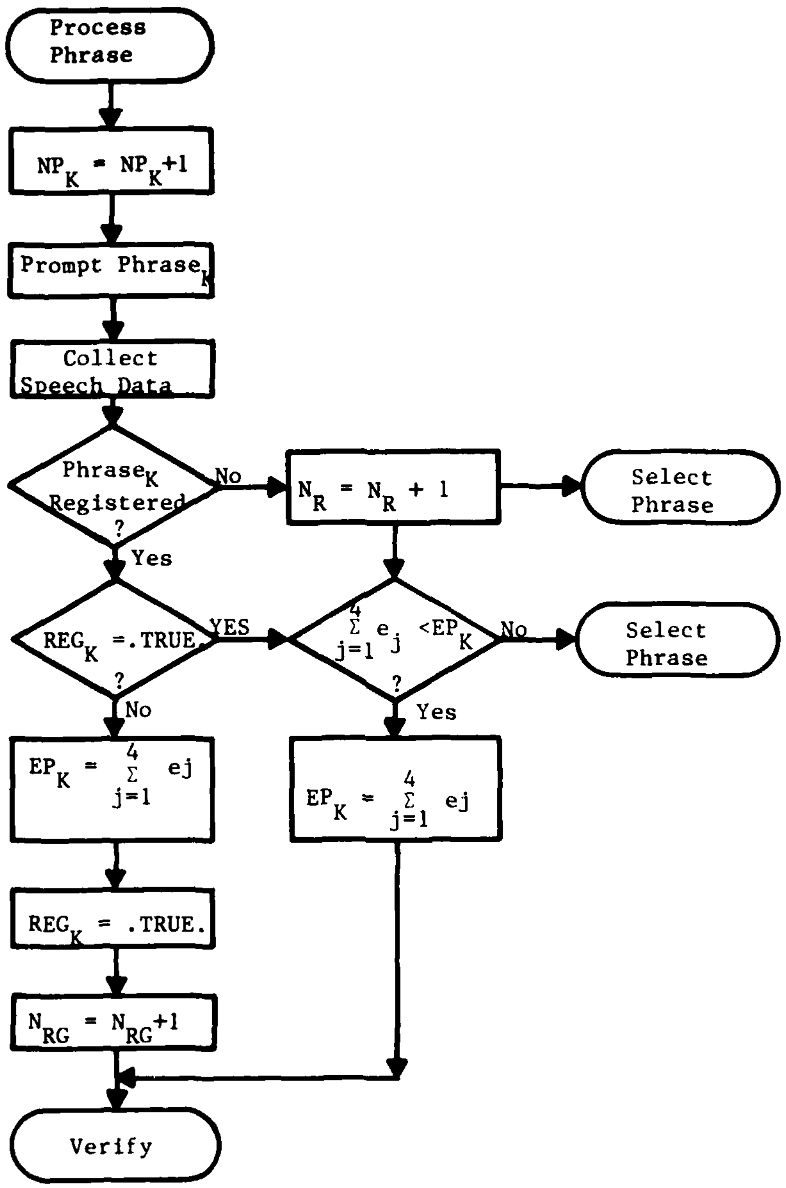
USED$_i$ = .TRUE. if phrase i used for decision function computation

N = Number of phrases used in computing decision function

d_T = Decision threshold

D_{T_N} = Decision thresholds as a function of # of phrases used

D_{TP_N} = Prejudiced decision thresholds as function of N



EP_K = sum of valley point errors for $K^{\underline{th}}$ phrase

EXP_K = expected sum of valley point errors for $K^{\underline{th}}$ phrase

where the best (4 or less), registered phrases are used in the calculation of the decision function. If the user is not verified, the worst phrase, as before, is excluded and the decision function is recalculated. This procedure is repeated until either the user has been verified or all phrases have been excluded. Then, if the user has not been verified, less than seven phrases have been prompted, and no more than three phrases have been misregistered, another phrase is prompted and the decision procedure is repeated.

3. DECISION FUNCTIONS

The decision function that was used for both the BISS and the CIC systems was of the form:

$$D = \frac{\sum_{k=1}^{N} \sum_{i=1}^{4} e_{ik}}{\max\left[\min\left(\sum_{k=1}^{N} \sum_{i=1}^{4} \hat{e}_{ik}\ ,\ 4N\hat{e}_{max}\right),\ 4N\hat{e}_{min}\right]} ,$$

where $\hat{e}_{max} = 140$ and $\hat{e}_{min} = 100$ for BISS, but varies by phrase for CIC. However, it was felt that impostor rejection on 1-phrase decisions could be improved by computing the min/max limits on error normalization over the entire reference file rather than individually on each subset of phrases. Hence, a new decision function was defined as

$$D' = \frac{\sum_{k=1}^{N} \sum_{i=1}^{4} e_{ik}}{\sum_{k=1}^{N} \sum_{i=1}^{4} \hat{e}_{ik}}$$

where

$$\beta = \frac{\sum_{k=1}^{4} \sum_{i=1}^{4} \hat{e}_{ik}}{\max\left[\min\left(\sum_{k=1}^{4} \sum_{i=1}^{4} \hat{e}_{ik}\ ,\ 16\hat{e}_{max}\right),\ 16\hat{e}_{min}\right]}$$

(For a 4-phrase decision, D′ is identical to D.)

4. DECISION PARAMETERS

Since different preprocessing methods yield different ranges of "scanning errors," it became necessary to adjust the various parameters in the decision function equation. In addition, the desire to improve the selectivity of the "good" phrases, precipitated decision parameters that were a function of the number of phrases in the decision function. Hence, there became more parameter variations than for BISS, as shown in Tables 20 and 21.

TABLE 20. SETS OF DECISION FUNCTION THRESHOLDS USED

	BISS	Set 1		Set 2		Set 3
Post-enrollment:						
Normal & Prejudiced** Mode:						
1-3 Phrases:	0	0		0		0
4 Phrases :	145	135	----->	145		145
Post-post-enrollment:						
Normal & Prejudiced Mode:						
1 Phrase:	*	85		85		85
2 Phrases:	*	110		110		110
3 Phrases:	*	125	----->	120		120
4 Phrases:	*	135	----->	140		140
Normal:						
Normal Mode:						
1 Phrase:	100	105		105	----->	100
2 Phrases:	120	125		125		125
3 Phrases:	135	130		130		130
4 Phrases:	145	135		135		135
Prejudiced Mode**:						
1 Phrase:	85	85		85		85
2 Phrases:	110	110		110		110
3 Phrases:	130	120		120		120
4 Phrases:	145	130		130		130

* "Post-post-enrollment" did not exist for BISS.
** Prejudiced mode parameters are "auto-abort" parameters for BISS

TABLE 21. SETS OF MAX/MIN SCANNING ERRORS ALLOWED (EMAX/EMIN)

	BISS	Set 1	Set 2	Set 3
Post-enrollment:				
Emax: 1-4 Phrases:	140	155	186	175
Emin: 1-4 Phrases:	100	120	144	130
Post-post-enrollment:				
Emax: 1 Phrase:	*	120	144	130
2 Phrases:	*	130	156	145
3 Phrases:	*	140	168	160
4 Phrases:	*	150	180	175
Emin: 1 Phrase:	*	100	120	100
2 Phrases:	*	105	126	110
3 Phrases:	*	110	132	120
4 Phrases:	*	115	138	130
Normal Mode:				
Emax: 1 Phrase:	140	130	156	145
2 Phrases:	140	135	162	155
3 Phrases:	140	140	168	165
4 Phrases:	140	145	174	175
Emin: 1-4 Phrases:	100	100	120	115

* "Post-post-enrollment" did not exist for BISS.

D. OTHER MODIFICATIONS

There were two significant modifications other than the ones described in this section. The first was the inclusion of a "post-post-enrollment" mode for users who have had from 4 to 12 previous verifications. The "normal" mode then became the mode for users when they have had twelve or more verifications. These mode designations affect the choice of decision thresholds and parameters, as explained earlier, and the choice of updating factors for reference files. The scheme for updating is to add 1/A times each spectral element of the "scanning" pattern formatted from the input speech to (1-A)/A times each spectral element of the reference scanning pattern. With the addition of the new mode, the updating factors, A, become A=4 for "post-enrollment" mode (0-3 prior verifications), A=8 for "post-post-enrollment" mode (4-11 prior verifications), and A=16 for "normal" mode (>11 prior verifications).

The other algorithm modification was the inclusion of numerous criteria to be used for terminating enrollment. (The BISS specification included in Appendix III uses "terminating enrollment" for successful completion of enrollment and "interruption" for the unsuccessful completion of enrollment. In the DSG/VVU system, the corressponding terms are "completion" and "termination," respectively. This discussion will use the DSG/VVU terminology.) For BISS, the only method of termination is by operator intervention. For the old 980-based CIC system, the following criteria were added for the termination of enrollment:

1. There is no speech input for <u>three</u> consecutive prompts.

2. The average scanning error (difference between input and reference) across all words exceeds 160 for preprocessing using Legendre polynomial regression, or exceeds 200 for preprocessing using sine/cosine regression.

3. The number of times required in establishing or reestablishing reference point data exceeds 11.

4. The reestablishment of reference points is required for a phrase (necessary when the same phrase does not register two consecutive times) after twenty phrases in all have been prompted.

For the DSG/VVU system, the following criteria were added to those used in the old CIC system:

1. There is no speech input for <u>two</u> consecutive prompts.

2. The total number of prompts during the establishment of the initial reference point data exceeds seven.

3. The number of reprompts of the same phrase during the establishment of the initial reference point data exceeds three.

4. The number of reprompts needed to reestablish reference point data for a phrase exceeds four.

SECTION V

ALGORITHM MODIFICATION TESTING

As shown in Figure 1 (Section I), the 980-based entry control system installed at the Corporate Information Center at Texas Instruments provided an excellent vehicle for testing the various algorithm modifications described in Section IV. However, because of intersession variation of the speakers (colds, maturity in system use, etc.), it soon became evident that a more controlled set of repeatable data was needed to perform algorithm tradeoffs. To this end, an enrollment session and 21 verification sessions were collected from 11 speakers (6 men, 5 women) for use in these tradeoff studies. This section first describes experiments performed with the off-line data base, then those with the operational entry control system.

A. OFF-LINE 11-SPEAKER TESTING

1. THE DATA SET

The data used for all the off-line experiments in this section were used as input to a variety of simulated filter banks and consisted of one enrollment session of 48 phrases (two repetitions of a set of 24 phrases) and 21 verification sessions of 12 phrases (three repetitions of a set of four phrases) collected from eleven speakers (6 males, 5 females) over a period of about one month (no more than two sessions per day). An enrollment without any problems requires 20 phrases. The amount collected provides one extra set of 4 phrases, and an extra repetition for each phrase. The data were digitized at 12.5 kHz and then filtered using a variety of simulated filter banks. The same data were used in the LPC residual energy experiments described in Section VII of this report.

For one-phrase decisions, these data provided 2772 (11*12*21) trials. Since there were only eleven speakers, however, these are not 2772 independent trials.

The last four verification sessions (18 through 21) for each talker were used as input for the impostor trials. Since there were 6 males and 5 females, this yielded 200 (4*6*5 + 4*5*4) sessions, or 2400 trials if each of the 12 phrases in each session were used in a one-phrase decision. All the experiments reported in this section were for one-phrase decisions.

2. FILTER BANK DEFINITIONS

This subsection provides a continuation of Table 10 given in Section IV for the various filter banks. The first column pair in Table 22 again shows the design center frequencies for the two-stage digital filters used for the BISS ADM. (Center frequencies of each of the two stages were slightly offset.) The center frequencies and bandwidths for

the "Total Voice" work are shown in the second column pair.

The set of simulated filter banks used in the experiments in this section were an outgrowth of the Total Voice filters, and hence were referred to as TV prime (TVP), with the bandwidth of the initial filter following (e.g., TVP200). These experiments showed that speaker verification performance improved as the bandwidths of the filters decreased. The result was that the 14-channel DSG filter bank, was selected to be defined as the TVP filter bank having an initial filter bandwidth of 200 Hz. The iterative definition of these filter banks is again given by

$$BW_{n+1} = BW_n (2)^{1/13} \qquad CF_{n+1} = CF_n + (CF_n - CF_{n-1}) (2)^{1/6}$$

where, $BW_1 = 200$, $CF_1 = 350$, and $CF_0 = 350 - 100/(2)^{1/6}$

Shown also are bandwidths with $BW_1 = 300$, $BW_1 = 250$, and $BW_1 = 150$.

TABLE 22. FILTER BANK DEFINITIONS

	BISS ADM (DESIGNED)		TOTAL VOICE (DESIGNED)		TVP (DSG/VVU)				
FILTER	C.F.	BW.	C.F.	BW.	C.F.	BW.	BW.	BW.	BW.
1	410	290	350	300	350	200	300	250	150
2	575	290	450	300	450	211	307	259	162
3	741	290	555	310	562	223	314	269	174
4	906	290	670	340	688	235	321	279	188
5	1071	290	790	380	830	248	328	289	203
6	1237	290	940	400	988	261	335	300	219
7	1402	290	1120	400	1167	275	343	311	236
8	1567	290	1320	400	1367	290	350	322	254
9	1733	306	1550	400	1591	306	358	334	274
10	1898	290	1810	400	1843	323	366	346	296
11	2063	290	2100	400	2126	341	374	359	319
12	2229	290	2430	400	2443	360	383	372	344
13	2394	290	2800	400	2800	379	391	386	371
14	2559	290	3350	700	3200	400	400	400	400
15	*2725	290	4000	700	**3649	422	409	415	431
16	*2890	290	4650	700	***4153	445	418	430	465

* Not used in ADM, although implemented in actual filters
** As defined for CCD analyzer; does not exist in DSG filter board.
*** Cut-off freq for high pass filter as defined for CCD analyzer; does not exist on DSG board.

3. PERFORMANCE MEASURES

The traditional error measure in measuring speaker verification performance has been the equal error point, i.e., the percent error at the acceptance threshold where the Type I (true speaker rejection) and the Type II (impostor acceptance) error rates are equal. Sometimes, however, it is not the equal error rate that is of interest, but the percent of one type of error given an acceptable value for the other error. An example of this is the BISS desire for the lowest Type II error rate given a Type I error rate of 1%. This complicates the comparison of systems that have reported their error rates in different ways. The observation that the product of the Type I error rate and the Type II error rate at a given threshold is relatively constant for all thresholds has prompted Texas Instruments to begin measuring performance with the use of a probability product, as defined by

$$E_{rr} = \sqrt{\frac{1}{N_{TS} * N_{IM}} * \sum_{k=1}^{\infty} N_{TS_k} * N_{IM_k}} ,$$

where

N_{TS} = total number of true speaker trials,

N_{IM} = total number of impostor trials,

N_{TS_k} = number of true speaker errors at threshold "k",

N_{IM_k} = number of impostor errors at threshold "k".

In fact, for smaller experiments, a smoothing of the product has found to be desirable, as defined by

$$\bar{E}_{rr} = \sqrt{\frac{1}{N_{TS} * N_{IM}} * \sum_{k=2}^{\infty} \left(\frac{1}{3} \sum_{j=k-1}^{k+1} N_{TS_j} * \frac{1}{3} \sum_{j=k-1}^{k+1} N_{IM_j} \right)} ,$$

where the parameters are as defined above. Both of these error measures were used in this chapter to evaluate performance of the various experiments.

4. RESULTS

The results of the experiments run on the 11-speaker data base are shown in Tables 23 and 24. Table 23 shows the expected (scanning) er-

Table 23

Speaker Expected Errors

Speaker	Enrollment					
	BISS	IPMOD2	TVP(300)	TVP250	TVP200	TVP150
BS	142	156	152	148	149	152
CC	88	118	102	93	95	100
CW	161	185	192	194	190	196
TK	83	107	92	85	85	87
GD	85	108	107	96	98	98
JH	92	108	99	88	84	91
Male Average	109	130	124	117	117	121
(Male σ)	(34)	(33)	(40)	(44)	(43)	(44)
BH	118	154	149	130	131	128
DD	100	145	119	111	111	117
FG	116	119	115	111	113	124
YH	142	159	156	183	182	199
LM	131	160	142	133	140	129
Female Average	121	148	136	134	135	139
(Female σ)	(16)	(17)	(18)	(29)	(29)	(34)
Average	114	138	130	125	125	129
(σ)	(27)	(27)	(31)	(37)	(37)	(39)

Speaker	After 21 Sessions					
	BISS	IPMOD2	TVP(300)	TVP250	TVP200	TVP150
BS	104	117	123	116	114	116
CC	77	103	90	81	82	82
CW	124	136	157	142	142	145
TK	75	97	102	90	88	91
GD	88	104	113	102	101	105
JH	79	95	91	84	81	83
Male Average	91	109	113	103	101	104
(Male σ)	(19)	(15)	(25)	(23)	(24)	(24)
BH	111	124	138	126	132	135
DD	88	119	95	92	93	99
FG	106	117	113	107	111	123
YH	107	131	109	124	131	142
LM	112	138	133	123	125	129
Female Average	105	126	118	114	118	126
(Female σ)	(10)	(9)	(18)	(15)	(16)	(16)
Average	97	116	115	108	109	114
(σ)	(17)	(15)	(21)	(20)	(22)	(23)

Table 24

Performance Comparison of Processing Methods

Processing Method	$\bar{E}_{rr}:E_{rr}$:# Errors at Equal Error Point:Threshold			
	D = $\bar{e}$	D = $\bar{e}/\hat{e}$	D = CIC	D = CIC'
BISS				
Male	: : 26:170	: :35:1.61	: :24:1.51	
Female	: : 76:162	: :38:1.54	: :47:1.45	
Total	: :104:164	: :75:1.57	: :71:1.47	
IPMOD2				
Male	:0.64: 21:200	:0.74:23:1.65	:0.60:18:1.46	:0.56:17:1.53
Female	:1.91: 58:192	:1.13:30:1.51	:1.41:34:1.44	:1.13:28:1.48
Total	:1.70: 77:195	:0.93:59:1.56	:0.95:52:1.45	:0.82:45:1.50
TVP (300)				
Male	: : 30:202	: :29:1.66	: :22:1.58	
Female	: : 41:205	: :23:1.59	: :23:1.54	
Total	: : 72:204	: :52:1.63	: :43:1.56	
TVP250				
Male	0.76:0.83: 25:188	0.72:0.73:26:1.65	0.55:0.57:16:1.51	0.47:0.47:17:1.52
Female	1.17:1.19: 37:199	0.76:0.77:21:1.62	0.87:0.87:23:1.56	0.71:0.74:21:1.58
Total	0.94:0.98: 62:194	0.74:0.75:47:1.65	0.70:0.71:39:1.54	0.58:0.57:40:1.55
TVP200				
Male	0.80:0.79: 25:188	0.79:0.75:29:1.64	0.66:0.61:19:1.45	0.59:0.53:16:1.46
Female	1.05:1.08: 32:203	0.70:0.73:18:1.60	0.75:0.74:18:1.59	0.65:0.66:16:1.57
Total	0.93:0.93: 54:196	0.74:0.74:46:1.62	0.72:0.70:38:1.52	0.63:0.59:36:1.51
TVP150				
Male	0.74:0.77: 26:194	0.81:0.79:30:1.64	0.62:0.64:19:1.47	0.55:0.53:17:1.45
Female	0.74:0.76: 23:225	0.55:0.51:16:1.70	0.56:0.54:14:1.62	0.51:0.52:13:1.64
Total	0.85:0.86: 52:207	0.68:0.67:45:1.65	0.64:0.64:33:1.54	0.59:0.57:33:1.54

rors for all speakers, both after enrollment and after 21 sessions, for all experimental conditions. Table 24 shows the verification results for up to four different decision functions: the average error, the normalized average error, and the two decision functions (D and D′) defined in Section IV. Up to four numbers are shown for each experiment: the average (smoothed) probability product, the raw probability product, the number of errors at the equal error rate, and the threshold at the equal error rate. Note that there were 1512 male true speaker trials, 1260 female true speaker trials, 1440 male impostor trials, and 960 female impostor trials.

The first two experiments, labeled BISS and IPMOD2 in the tables, were both run using the simulated BISS filter bank to test for any difference caused by the preprocessings. The other four experiments were run using the simulated TVP filter banks, with varying bandwidths. The first experiment was run using the BISS type preprocessing. All of the other experiments were run using the DSG (IPMOD2) type preprocessing.

Comparison of the first two experiments clearly shows the results for all three decision functions in Table 24 are better using the DSG (IPMOD2) preprocessing.

Comparison of the second and third experiments shows that the male performance degrades slightly, while the female performance improves dramatically. Although the overall error rate decreases for the TVP filters, more importantly, the error rate becomes more uniform across sex.

The performance also improves for the TVP filters as the bandwidth of the bandpass filters is decreased. Since it was believed that the bandwidth for the TVP150 set of filters left too many spectral holes and that its lower error rate might be an artifact caused by too small a sample size, the TVP200 filter bank was chosen for the DSG/VVU systems.

Performance clearly improves from left to right in **Table 24.** More importantly, it is the female performance that most drastically improves between the first column (unnormalized decision function) and the other three columns (all normalized decision functions).

B. ON-LINE OPERATIONAL TESTING FOR 980-BASED CIC SYSTEM

1. THE DATA

The data used to derive the true speaker (Type I) performance were the same data used during the normal operation of the booth. Almost none of these data were actually recorded, as they were during the VVU operational test. The statistics for these tests were those derived from the results stored for every access attempt. The information recorded for every entry attempt consisted of data such as times and values of the scanning errors at the selected reference point location, expected values of scanning errors for the user for that trial, trial

number, etc.

In addition to the data used during the normal operation of the booth, two sets of data were collected on-line in the CIC entry control booth for performing off-line Type II trials against the actual CIC reference files. The data collected were the preprocessed data (filtered, regressed, normalized, and quantized) from specially collected sessions of 12 phrases each (three repetitions of four phrases). Since the CIC system was running with both types of preprocessings (BISS type and DSG or IPMOD2 type), data were collected from both types of users. The number of impostor sessions were 40 (all males) for the BISS-type preprocessing and 32 (23 males, 9 females) for the DSG-type (IPMOD2) preprocessing. The number of references varied, depending on the number available when a particular experiment was performed.

2. SYSTEM MODIFICATION LOG

The log in Table 25 is given to correlate the modifications to the system with any changes in performance. When the DSG-type preprocessing was installed, a common verification algorithm was used, but two separate preprocessing routines and some parameters were maintained. Also, after this time, all enrollments were performed with the new preprocessing; hence the performance for the BISS-type preprocessing would naturally improve due to the increasing maturity of the user population when no new users were being enrolled.

Note that references to different decision functions, different energy definitions, and various sets of parameters may be resolved by referring to the description of the algorithm modifications in Section IV of this report.

3. RESULTS OF ON-LINE OPERATIONAL TESTING

The results of both the on-line testing with the operational CIC system and the off-line impostor testing are shown in Tables 26 and 27, for the BISS and DSG (IPMOD2) preprocessings, respectively. In both tables the test conditions are given at the top with both true speaker and impostor test results for a given set of conditions given below. Notable in these tables is the lack of performance improvement as the result of any of the parametric modifications. The only significant change for the true speaker results is the shift of some of the verifieds from taking one phrase to taking two phrases when the one-phrase, normal mode decision threshold is changed from 105 to 100 (changing from decision threshold set 2 to set 3). The impostor testing results, however, show that the impostor acceptance rate when using the DSG (IPMOD2) preprocessing is about half of the impostor acceptance rate when using the BISS preprocessing. There is probably even more difference than shown in the tables of results, because the BISS preprocessing Type II results are for only males, and the results would be worse if both males and females were used in the Type II tests.

Finally, the impostor data using just the single-phrase average scanning error were combined with single-phrase average scanning errors

collected from the actual CIC system, and was used to plot Type I/Type II curves for a one-phrase strategy for both BISS and IPMOD2 types of preprocessing. To obtain plots with comparable abscissas, the IPMOD2 average scanning errors were multiplied by "R," where "R" is the ratio of the average true speaker scanning errors for IPMOD2 preprocessing to that for BISS preprocessing. A plot of the average scanning errors for single phrases are shown in Figure 20, where "R" is 1.25.

TABLE 25. LOG OF CHANGES TO THE 980-BASED CIC SYSTEM

Date	BISS PREPROCESSING	DSG PREPROCESSING
spring 77	7-phrase dec. strat. installed	
11/ /77	Statistics collection begun	
7/12/78	Set 1 decn threshld installed	
wk.10/16/78	1. Max VPE from 600 to 400 2. proc. cut-off delta from 5 to 20. (only suggested??)	
11/18/78	Set 2 dec. thrshlds installed	
03/13/79	D´ installed	DSG prepro. installed with D´ & max ehat after enroll = 1.2 * BISS parameters. Set 2 dec thrshlds Set 2 emax/emin´s
04/05/79	Energy def. 2 Energy PVR from 10 to 20 Min EOS Energy from 70 to 40 No. EOS Samples from 50 to 70	Energy def. 2 Energy PVR from 10 to 20 Min EOS Ener. frm 70 to 40 No. EOS Smpls frm 50 to 70 Max VPE frm 400 to 480
wk.06/03/79	Max. No. same phr reprompts for PE & Norm from 2 to 1	Max. No. same phr reprompts for PE & Norm from 2 to 1
wk.07/09/79	Normal 1 phrase dec thrsh. from 105 to 100	Normal 1 phrase dec thrsh. from 105 to 100 Set 3 emax/emin´s max ehat after enroll from 192 to 200 Max VPE from 480 to 500
05/18/80	(980 ASV system removed from CIC)	

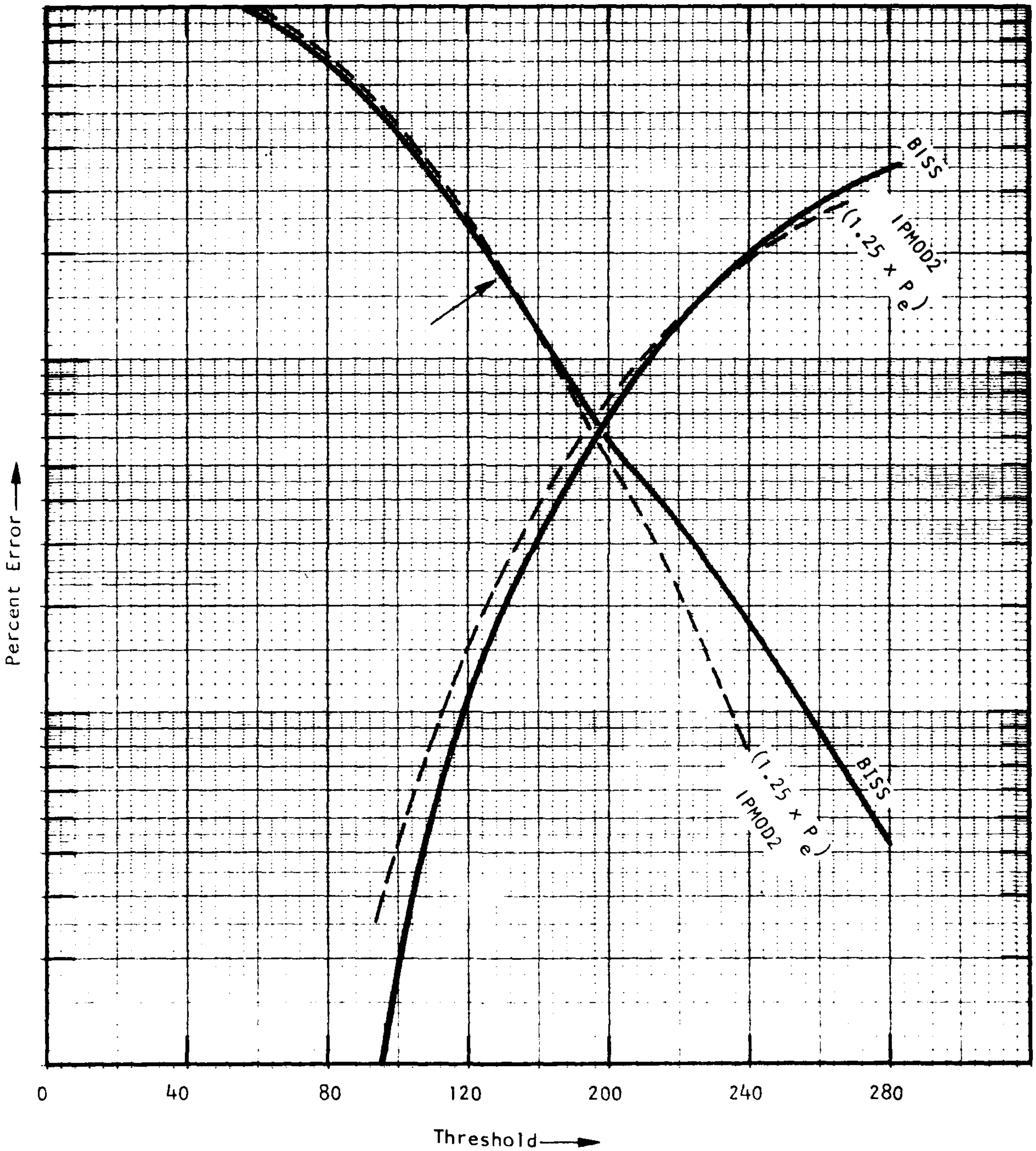


Figure 20 Comparison of BISS/IPMOD2 Single-Phrase Performance for CIC

TABLE 26. ON-LINE TRUE SPEAKER AND OFF-LINE IMPOSTOR RESULTS FOR CIC ENTRY CONTROL BOOTH USING BISS-TYPE PREPROCESSING

Test Conditions:							
Dec´n Fcn:	D	D	D ----->	D´	D´	D´	D´
Max VPE:	600 --->	400	400	400	400	400	400
Dec´n Thres:	Set 1	Set 1--->	Set 2	Set 2	Set 2	Set 2--->	Set 3
Energy Def:	Def 1	Def 1	Def 1	Def 1--->	Def 2	Def 2	Def 2
Energy PVR:	10	10	10	10 --->	20	20	20
Min EOS En:	70	70	70	70 --->	40	40	40
# Smpls EOS:	50	50	50	50 --->	70	70	70
Max # Same Phr Reprompts:							
PE:	2	2	2	2	2	2	2
PPE/Norm:	2	2	2	2	2 ----->	1	1
Enrollments:	yes	yes	yes --->	no	no	no	no

On-line True Speaker Trials:							
Srt Date:	07/18/78	10/24/78	11/21/78	03/13/79	04/07/79	06/05/79	07/15/79
End Date:	10/24/78	11/21/78	03/12/79	04/04/79	06/05/79	07/10/79	05/18/80
# Trials:	52,419	14,458	48,759	9,938	23,360	14,851	91,266
Not Verf´d:	0.51%	0.59%	0.67%	0.78%	0.62%	0.45%	0.35%
	(265)	(86)	(328)	(78)	(145)	(67)	(316)
No Respnse:	0.20%	0.10%	0.25%	0.16%	0.28%	0.11%	0.10%
	(104)	(14)	(122)	(16)	(66)	(16)	(93)
% Verf´d on phrase 1:	74.35%	72.55%	71.80%	74.14%	74.13%	75.52%	70.21%
2:	18.82	19.30	20.74	19.24	19.52	18.78	23.77
3:	4.05	4.50	4.55	3.80	3.96	3.64	3.86
4:	1.75	1.66	1.80	1.80	1.43	1.28	1.34
5:	0.59	0.86	0.69	0.58	0.55	0.42	0.49
6:	0.31	0.32	0.26	0.28	0.28	0.25	0.21
7:	0.13	0.12	0.16	0.16	0.13	0.11	0.11

Off-line Impostor Trials (10320 trials):				
Test Date:	09/18/78	11/13/78	02/05/79	
Max VPE:	600	600	600	400
Imp. Accepts:				
Overall:	1.30%	1.38%	1.28%	1.25%
	(134)	(142)	(132)	(129)
Normal Mode:				
1 Phrasse:	36	36	33	33
2 Phrases:	22	22	23	22
3 Phrases:	14	14	16	14
4 Phrases:	15	23	21	17
Prej. Mode:				
1 Phrase:	10	9	2	2
2 Phrases:	11	12	14	18
3 Phrases:	12	10	6	6
4 Phrases:	14	16	17	17

TABLE 27. ON-LINE TRUE SPEAKER AND OFF-LINE IMPOSTOR RESULTS FOR CIC ENTRY CONTROL BOOTH USING IPMOD2-TYPE PREPROCESSING

Test Conditions:					
Dec´n Function:	D´	D´	D´	D´	D´
Max VPE (TS trials):	400 --->	480	480 --->	500	
Max VPE (IM trials):			480	?	?
Dec´n Thres:	Set 2	Set 2	Set 2--->	Set 3	Set 3
Energy Def:	Def 1--->	Def 2	Def 2	Def 2	Def 2
Emax/Emin:	Set 1	Set 1	Set 1--->	Set 2	Set 2
Energy PVR:	10 --->	20	20	20	20
Min EOS En:	70 --->	40	40	40	40
# Smpls EOS:	50 --->	70	70	70	70
Max # Same Phr Reprompts:					
PE:	2	2	2	2	2
PPE/Norm:	2	2 ----->	1	1	1
Enrollments:	yes	yes	yes	yes	
# male/female IM´s:			16/7	16/7 -->	23/9
# male/female ref´s:			50/9	50/9	?
On-line True Speaker Trials:					
Srt Date:	03/13/79	04/07/79	06/05/79	07/15/79	
End Date:	04/04/79	06/05/79	07/10/79	05/18/80	
# Trials:	191	2,902	3,288	44,406	
Not Verf´d:	3.14% (6)	0.69% (20)	0.49% (16)	0.70% (312)	
No Respnse:	0.16% (16)	0.28% (66)	0.11% (16)	0.10% (93)	
% Verf´d on phrase 1:	57.38%	57.38%	66.17%	59.80%	
2:	21.86	25.95	23.52	30.04	
3:	4.92	7.52	5.30	5.84	
4:	11.48	6.11	3.09	2.77	
5:	1.64	1.63	1.19	0.97	
6:	1.64	1.11	0.55	0.44	
7:	1.09	0.28	0.18	0.15	
Off-line Impostor Trials (10320 trials):					
Test Date:			07/10/79	07/10/79	08/08/79
# Trials:			857	857	1,223
Imp. Accepts:					
Overall:			1.28% (11)	0.82% (7)	0.65% (8)
Normal Mode:					
1 Phrase:			6	3	3
2 Phrases:			1	1	1
3 Phrases:			0	0	0
4 Phrases:			0	0	0
Prej. Mode:					
1 Phrase:			1	0	0
2 Phrases:			0	1	2
3 Phrases:			0	0	0
4 Phrases:			3	2	2

SECTION VI

TEST OF THE VOICE VERIFICATION UPGRADE SYSTEM

Enrollment of users on the Voice Verification Upgrade (VVU) system installed to control one of the entries to the Semiconductor Building at Texas Instruments in Dallas, began on 12 December 1980. Operation on a 24-hour per day basis began on 2 January 1981 and continued through 17 June 1981. New users were enrolled on the system throughout its operation, for a total of 286 users (200 men, 86 women). These users produced 13,539 trials (presumably all "true-speaker" trials), with a true speaker rejection rate of 2.26% overall and 1.04% during the normal mode of operation (i.e., more than 11 prior verifications).

During normal usage of the system, updated reference files were selectively captured for later use in casual impostor trials. The reference files produced after enrollment and after 1,2,4,8,16,32,64, etc. previous verifications were the reference files that were saved. During the end of May 1981, one special impostor session was collected from each of 75 users (51 men, 24 women) who volunteered. Each impostor session was collected in the same entry control booth with the same type of protocol. The only difference was that twelve phrases (three repetitions of a set of four phrases) were required rather than the usual one or two phrases. The data were preprocessed and sent from the terminal processor to the host processor, where it was stored on disk.

The next two subsections describe the results of the normal booth usage (true speaker rejection rates) and the casual impostor testing results.

A. "TRUE SPEAKER" TESTING

The "true speaker" (assuming no casual impostor attempts by the users) testing was performed using the VVU system in an operational environment controlling an entrance to the Semiconductor Building at Texas Instruments. Since there were other entrances to this building controlled either by guard or by closed-circuit television, all users enrolled on the voice system had the option of using other entrances; however, the placement of the voice controlled entrance was chosen to provide the benefit of convenience to many of those using the entrance. Although enrollments began on the system on 12 December 1980, verifications on a 24-hour per day basis did not begin until 2 January 1981. The system remained operational except for occasional maintenance until its delivery to RADC on 17 June 1981.

The true speaker testing yielded a higher percentage of not-verifieds than was expected. However, all users of the system were inexperienced in the use of such a system, with no large population of "veteran" users such as existed when data were first collected either from the old 980-based CIC system or from the newer 990-based CIC system. (Most users of the newer system had been enrolled on the old CIC system.) The gross rejection rate from 2 January 1981 through 17 June

1981 was 2.26% (306 rejections out of 13,539 total trials) by 247 users (194 males, 64 females). (39 of the 286 users never used the system after enrolling.) The improved performance as a function of increasing experience of the user population can be seen by comparing the gross rejection rate of 0.69% (5 rejections out of 719 total trials) for the last two weeks of operation of the booth, to the 2.26% for the entire testing period.

Typically, eighteen (7.3%) of the users accounted for 156 (51%) of the errors. Table 28 shows the distribution of the number of not-verifieds across the user population.

TABLE 28. NOT VERIFIEDS ACROSS THE USER POPULATION

# OF N.V.	# OF USERS	# OF N.V.	# OF USERS
0	144	8	2
1	43	9	5
2	25	10	1
3	11	11	0
4	6	12	2
5	4	13	0
6	1	14	2
7	1		

The importance of the reenrollment capability is demonstrated by the dramatic performance improvement shown in Table 29 for the eight users who were reenrolled.

TABLE 29. PERFORMANCE IMPROVEMENT FOR EIGHT REENROLLED SPEAKERS

USER	BEFORE REENROLLMENT #N.V./#USES	AFTER REENROLLMENT #N.V./#USES
209230	8/8	4/140
598247	3/4	0/31
674419	8/43	2/33
1762517	3/3	0/3
3452604	13/16	1/3
9540423	2/6	0/24
12462163	9/9	0/29
14020636	4/5	5/60
TOTALS	50/94 (53.2%)	12/323 (3.7%)

When the statistics before reenrollment for these eight users are excluded, the overall rejection rate becomes 1.95% (256 rejections out of 13,139 uses).

Of special interest is the performance as a function of the number of prior verifications, which is shown in Table 30 for both the [illegible] VVU, 990-based systems. The rejection rates shown in this table [illegible] 16.15% (12.62% excluding data prior to reenrollment [illegible] "post-enrollment" (0-3 prior verifications) per[illegible] 4[illegible] "post-post-enrollment" (4-11 prior verifications [illegible]

AD-A118 444 TEXAS INSTRUMENTS INC DALLAS F/G 17/2
VOICE VERIFICATION UPGRADE.(U)
JUN 82 R L DAVIS, J T SINNAMON, D L COX F30602-79-C-0190
UNCLASSIFIED TI-08-82-07 RADC-TR-82-139 NL

2 OF 3
AD
A118444

(0.96% excluding data prior to reenrollments) for the "normal" (more than 11 prior verifications) period.

TABLE 30. PERFORMANCE AS A FUNCTION OF NUMBER OF PRIOR VERIFICATIONS FOR THE 990-BASED SYSTEMS

	CIC (since 3/11/80)		VVU (since 1/2/81)	
	# trials	%	# trials	%
Verified:	43,418	99.56	13,233	97.74
Not-verified:	193	0.44	306	2.26

Reject rate vs # of prior sessions:	# trials	# Errors	% Error	# trials	# Errors	% Error
0- 3:	371		3.77	1,065	172	16.15
4- 11:	770		0.52	1,492	22	1.47
12- 27:	1,891		0.69	2,504	41	1.64
28- 59:	3,699		0.54	3,712	26	0.70
60- 123:	6,521		0.25	3,949	37	0.94
124- 251:	6,498		0.51	814	8	0.98
252- 507:	11,246		0.45	0	0	--
508-1019:	12,323		0.39	0	0	--
1020-2043:	283	0	0.00	0	0	--
2044-4091:	0	0	--	0	0	--
4092-8187:	0	0	--	0	0	--
>8187:	0	0	--	0	0	--

Also of interest is the number of retries for users who are not verified. The 306 not-verifieds occurred during 195 sessions. This means that although the overall reject rate was 2.26% per trial, the overall reject rate on a booth passage basis, which is the real number of interest for an entry control system, was only 0.63% (84 rejections out of 13,317 access attempts). Although the users were instructed to retry once after their first rejection, but to quit and call for assistance (an override) after two rejections, Table 31 shows that a significant number of the users did not follow instructions. Since there were only 306 not-verified's during the testing, those users who tried several times in succession added a significant negative bias to the overall user acceptance rate. In addition, the retry capability also affords the opportunity for casual impostor attempts by curious users, also adversely affecting the user acceptance rate. However, the necessity of at least a limited retry capability can be seen by noting that 71 percent of the users who tried more than once were finally verified (60 percent on the second try), which as explained above, improves the acceptance rate on a booth passage basis. Table 31 shows the access reject rate ([number who retry + cumulative number who quit] / [number of access attempts]) when the number of verification attempts is limited to a specified number of trials. The data in this table suggests that it might be prudent to limit the number of trials per user I.D., per passage, to two or three.

TABLE 31. ACCESS REJECT RATE (13,317 ACCESSES) AS A FUNCTION OF NUMBER OF RETRIES

VERIF ATTEMPT NUMBER	# OF VERIF ATTEMPTS	NUMBER VER'D	CUM. NUMBER VER'D	# NOT VER'D	# WHO RETRY	# WHO QUIT	CUM. # WHO QUIT	REJECT RATE IF ATTMPTS LIMITED
1	13,317	13,122	13,122	195	157	38	38	1.46%
2	157	94	13,216	63	35	28	66	0.76%
3	35	9	13,225	26	15	11	77	0.69%
4	15	5	13,230	10	8	2	79	0.65%
5	8	2	13,232	6	3	3	82	0.64%
6	3	-	13,232	3	2	1	83	0.64%
7	2	1	13,233	1	1	-	83	0.63%
8	1	-	13,233	1	1	-	83	0.63%
9	1	-	13,233	1	0	1	84	0.63%

Although these performance statistics are biased by the large number of new users (lack of any long-term users), the performance was still not as good as expected, based on prior results from both the 980-based and the 990-based systems at CIC. The reasons for the poorer performance are judged to be:

1) The poor quality of the LPC prompting phrases (Although this was also a problem on the new CIC system, a majority of the users had been enrolled previously on the old 980 system, which used high quality PCM prompting phrases, and hence, were familiar with the word set).

2) The extended time (2-3 weeks) between enrollment and initial use by the first speakers to be enrolled (due to delays in completion of all of the software).

3) Getting the subjects to stand close enough and speak up loudly enough (Again, many of the users of the newer CIC system were experienced due to their use of the old 980 system).

4) Poor enrollments (This is felt to be due both to the poor quality prompting and the newness of the task).

Although the performance statistics were generally not broken down by sex for the VVU system, a few of the overall statistics were calculated and are shown in Table 32.

TABLE 32. VVU PERFORMANCE BY SEX

	MALES	FEMALES
Number of users: (with one or more uses)	194	69
Number of uses:		
total:	10,271	3,268
ave. per user:	56	51
Not-verified's		
number:	154	152
percent:	1.50%	4.65%
Ave. Ehat at End of Test:	97	134

A summary of the verification performance as a function of several other parameters is shown in Table 33. Of special note are that 1) the average number of phrases for verification was 1.76, 2) the reject rate is higher with more than one person in the booth (3.3% vs. 2.0%), 3) the reject rate decreases with usage, and 4) the reject rate increases as the expected "error" (data variance) for the speaker increases.

B. "CASUAL" IMPOSTOR TESTING

The "casual" impostor testing was performed by using specially collected impostor sessions from users enrolled on the system against user reference files that had been stored during normal use of the system. The impostor files consisted of preprocessed (regressed, normalized and quantized) filter bank data for three repetitions of a set of four phrases. The data were collected on-line in the VVU booth by soliciting a voluntary impostor session from users during their normal use of the booth. Data for a total of 75 impostors (51 men, 24 women) were collected.

The reference data were collected automatically during normal booth usage by storing updated reference files after 0, 1, 2, 4, 8, 16, etc. previous verifications. Due to the quantity, these data were stored on the secondary disc drive, which on occasion was not on-line. Since the reference file data collection was peripheral to the actual booth operation, the VVU system would not cease operation if the secondary storage medium was off-line. Hence, occasionally (1% of the time) reference files were missed for some transactions. In addition, since some users were reenrolled, duplicate reference files for the same number of prior verifications were collected.

The actual impostor testing performed was to compare all impostors against all other users (of the same sex) for which reference files after 32 prior verifications existed. Of the 286 users (200 men, 86 women), only 116 of them (93 men, 33 women) completed 32 or more verifications. Impostor trials were also run against this set of users for their reference files immediately after enrollment. Since eight of the 51 male impostors and four of the 24 female impostors had not completed 32 sessions, this yielded 4,700 (43*92 + 8*93) independent trials for

Table 33

Verification Performance Summary (Type 1 Analysis) for VVU ASV System

DATA FOR 1 / 2 / 1981 AT 2043 THRU 6 / 17 / 1981 AT 1437

Table II(a)

```
# VERIFIED        = 13233 97.74%
  NOT VERIFIED =    306  2.26% : 306    2.26%  NV:    0    0.00%  NO RESPONSE
 TOTAL # PHRASES IN VERIFICATION = 23313 ;  AVG FOR VERIFICATION =     1.76
 TOTAL # PHRASES OVERALL         = 25343 ;  AVG OVERALL          =     1.87
 TOTAL # PHRASES NOT REGISTERED  =  1501 ;   ( 5.92% )
```

************NUMBER VERIFIED ON PHRASE PH*************

PH	*****ALL****** N	%	CUM%	*NONPREJUDICED* N	%	CUM%	**PREJUDICED*** N	%	CUM%	*NOT VERIFIED* N	%	CUM%
1	7266	53.7	53.7	7266	53.7	53.7	0	0.0	0.0	0	0.0	0.0
2	3825	28.3	81.9	3766	27.8	81.5	59	0.4	0.4	0	0.0	0.0
3	798	5.9	87.8	670	4.9	86.4	128	0.9	1.4	0	0.0	0.0
4	932	6.9	94.7	844	6.2	92.7	88	0.6	2.0	12	0.1	0.1
5	235	1.7	96.4	0	0.0	92.7	235	1.7	3.8	22	0.2	0.3
6	122	0.9	97.3	0	0.0	92.7	122	0.9	4.7	32	0.2	0.5
7	52	0.4	97.7	0	0.0	92.7	52	0.4	5.1	240	1.8	2.3

*****VERIFICATION STATISTICS*****

PH	# PHRASES USED N	%	CUM%	PHRASES DROPPED N	%	CUM%	# OF UNREG. PHRSES VERIFED NET	VERIFED TOTAL	NOT VERFD NET	NOT VERFD TOTAL	MAX # SAME-PHRSE REPROMPTS VERFD	NOT VRFD
0	0	0.0	0.0	12979	95.9	95.9	12944	12629	89	53	12823	12
1	7330	54.1	54.1	218	1.6	97.5	272	490	99	39	372	74
2	3899	28.8	82.9	27	0.2	97.7	13	90	71	35	35	220
3	759	5.6	88.5	6	0.0	97.7	1	21	34	69	0	0
4	1242	9.2	97.7	0	0.0	97.7	0	0	13	108	0	0

TRIALS VS # IN BOOTH:

#	*PASS* N	%	*VERFD* N	%	*NOT VERFD* N	%
1	0	0.0	10639	80.4	219	71.6
>1	0	0.0	2594	19.6	87	28.4

HISTOGRAM OF USAGE VERSUS TRIAL #:

TRIAL NUMBER	*TOTAL TRIALS* N	%	CUM%	*NOT VERIFIED* N	BIN%	CUM%
1 - 4	1065	7.9	7.9	172	16.2	16.2
5 - 12	1492	11.0	18.9	22	1.5	7.6
13 - 28	2504	18.5	37.4	41	1.6	4.6
29 - 60	3712	27.4	64.8	26	0.7	3.0
61 - 124	3949	29.2	94.0	37	0.9	2.3
125 - 252	814	6.0	100.0	8	1.0	2.3
253 - 508	0	0.0	100.0	0	0.0	2.3
509 - 1020	0	0.0	100.0	0	0.0	2.3
1021 - 2044	0	0.0	100.0	0	0.0	2.3
2045 - 4092	0	0.0	100.0	0	0.0	2.3
4093 - 8188	0	0.0	100.0	0	0.0	2.3
8189 - 32767	0	0.0	100.0	0	0.0	2.3

TRIALS VS TIME OF DAY:

TM	*PASS* N	%	*VER'D* N	%	*NOT VER'D* N	%
3	0	0.0	15	0.1	3	1.0
6	0	0.0	358	2.7	4	1.3
9	0	0.0	7039	53.2	149	48.7
12	0	0.0	1415	10.7	60	19.6
15	0	0.0	2321	17.5	63	20.6
18	0	0.0	1402	10.6	16	5.2
21	0	0.0	395	3.0	4	1.3
24	0	0.0	288	2.2	7	2.3

PHRASES UNREG'D ON PHRSE PH:

PH	N	PH%
1	388	2.9
2	254	4.1
3	236	9.7
4	181	11.0
5	203	28.9
6	161	36.1
7	78	26.7

HISTOGRAM OF USAGE VS SPKR AVG.:

AVG	*TOTAL TRIALS* N	%	CUM%	*NOT VER'D* N	BIN%	CUM%
0- 60	1190	8.8	8.8	9	0.8	0.8
61- 70	2277	16.8	25.6	24	1.1	1.0
71- 80	2966	21.9	47.5	53	1.8	1.3
81- 90	2908	21.5	69.0	51	1.8	1.5
91-100	1529	11.3	80.3	60	3.9	1.8
101-110	1470	10.9	91.1	56	3.8	2.1
111-999	1199	8.9	100.0	53	4.4	2.3

the men and 772 (20*32 + 4*33) for the women for the after 32-session reference files and somewhat less (as explained earlier) for the reference files immediately after enrollment. Hence, input from one speaker was tested against no more than two reference files for the same user.

The test against the reference files after 32 prior sessions used the parameters for the "normal" mode of verification and the test against the reference files after enrollment used the parameters for the "post-enrollment" mode of verification. The results of this casual impostor testing are given in Table 34. Of the 20 errors against enrollment reference files and the 44 errors against 32-session reference files, only nine of the errors were for the same impostor/reference pairs for the two types of reference files.

Also shown in Table 34 are the confidence levels that the true error is less than 1.0%, based upon the error rate observed in these experiments and assuming independent trials. For a fixed sample size, the confidence level decreases as the observed error rate approaches the desired upper bound on the error rate. This can be seen more graphically by the curve in Figure 21 that shows the upper limit on the observed error rate as a function of sample size necessary to insure a 90% confidence level that the true error rate is less than 1%. Also, for a fixed observed error rate, the confidence level increases as the sample size increases.

TABLE 34. CASUAL IMPOSTOR TESTING RESULTS AND CONFIDENCE LEVELS THAT TRUE ERROR RATE IS < 1.0% BASED UPON OBSERVED ERROR RATE

REFERENCE FILES AFTER	MEN #VER/#TRIALS	WOMEN #VER/#TRIALS	OVERALL #VER/#TRIALS
ENROLLMENT	19/4648 (0.41%)	1/795 (0.13%)	20/5443 (0.37%)
(Confidence level)	(99.9+%)	(99%)	(99.9+%)
32 VERIFICATIONS	43/4700 (0.91%)	1/772 (0.13%)	44/5472 (0.80%)
(Confidence level)	(69%)	(99%)	(92%)

The distribution of successful impostors and of references that were successfully impersonated was very nonuniform for the men. The same phenomenon would be expected for the women with a larger sample size. Those impostors successful against many users are known as "wolves," and those references who are successfully impersonated by many impostors are known as "eels." These distributions are shown in Table 35.

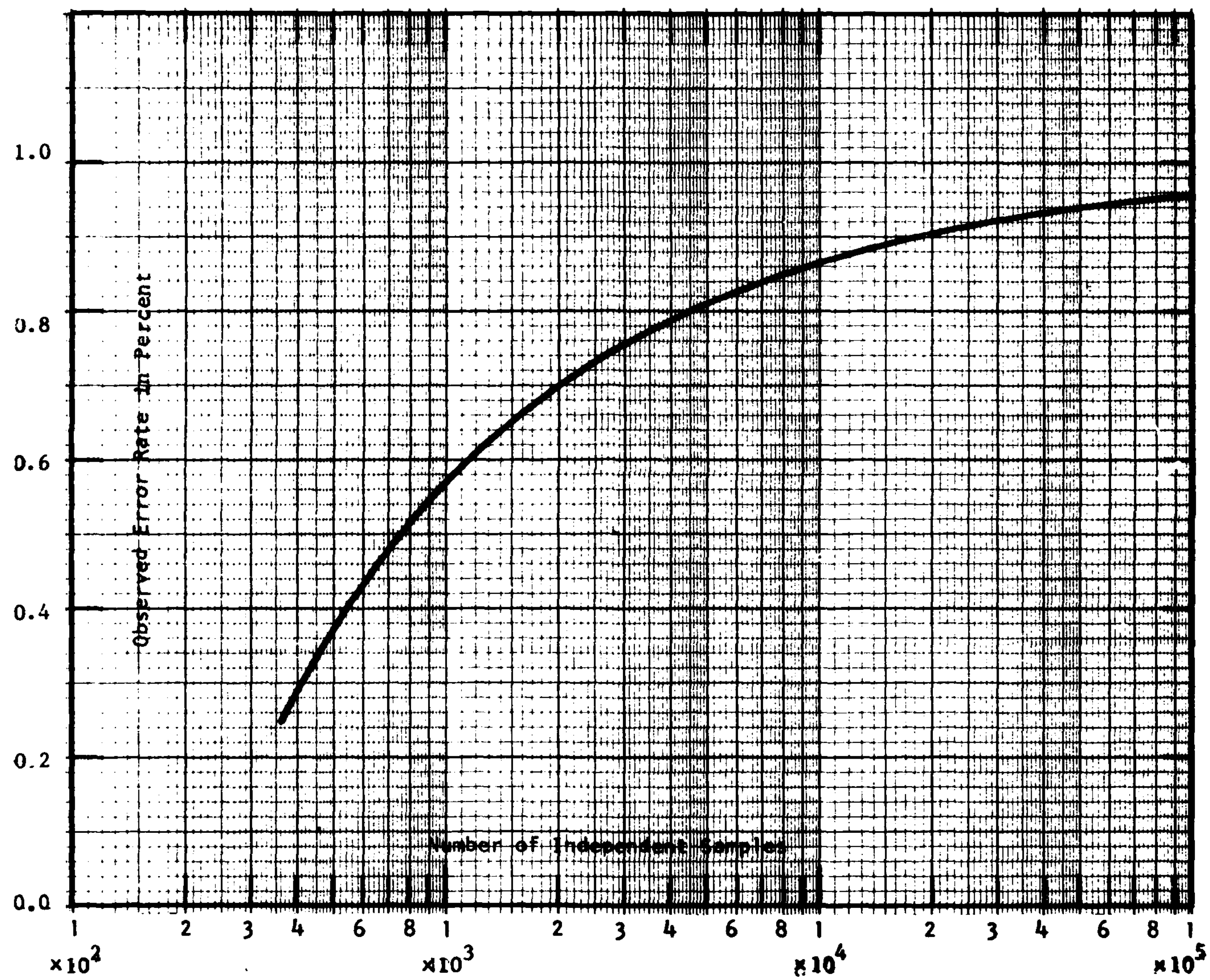


Figure 21 Observed Error Rate vs Sample Size to Insure a 90% Confidence Level that the True Error Rate ≤ 1%

TABLE 35. DISTRIBUTION OF IMPOSTOR SUCCESSES

# OF SUCCESSES	# OF IMPOSTORS			# OF REFERENCES		
	MEN	WOMEN	TOTAL	MEN	WOMEN	TOTAL
0	34	22	56	64	31	95
1	5	2	7	15	2	17
2	3	-	3	8	-	8
3	2	-	2	2	-	2
4	1	-	1	-	-	-
5	1	-	1	1	-	1
6	2	-	2	-	-	-
7	1	-	1	1	-	1
8	1	-	1	-	-	-
9	1	-	1	-	-	-
10	-	-	-	-	-	-
11	-	-	-	1	-	1

The existence of fixed input data to compare against variable reference files also allowed the observation that the adaptation of the reference files, and not just the adaptation of the speaker, decreased the number of phrases needed to verify (i.e. improved the performance). The average number of phrases to verify as a function of the number of prior verifications is given below for all successful verifications for all users. (For this experiment, the restriction was removed that required at least four phrases be prompted during the "post-enrollment" period.)

# OF PRIOR VERIFICATIONS:	0	1	2	4	8	16	32	64	128
AVERAGE # OF PHRASES TO VERIFY:	3.16	3.01	2.83	2.68	2.60	1.97	1.77	1.60	1.27

SECTION VII

RESIDUAL ENERGY BASED SPEAKER VERIFICATION

A variety of signal processing algorithms exist for representing the speech signal in terms of time varying parameters. These parameters are used in speech/speaker recognition for determining a measure of similarity. Examples of these signal processing transformations are the direct spectral measurement which uses either a bank of bandpass filters or a discrete Fourier transform, the cepstrum, or a set of suitable parameters of a linear predictive model. Selection of the parameter set depends to a considerable degree on performance and implementation considerations. It is generally agreed that the linear modeling techniques have performance comparable to or better than other techniques. Results summarized in an RADC report [14] indicate that the recognition error obtained for a speaker dependent word recognition system using the LPC-based recognition techniques is a factor of three times lower than the recognition error using the bandpass filtering technique. The linear predictive coding (LPC) based approach to speaker verification was chosen for investigation during this contract because of its performance and its ease of implementation in a system using a single-chip, digital signal processor, designed to provide high-speed multiply-accumulate operations.

A. SIMILARITY MEASURE

Once the parameters have been extracted from the speech waveform, a similarity measure must be computed between those parameters and a stored reference. The similarity measure used was patterned after that of Itakura, [15] which uses a normalized prediction residual. The LPC prediction residual energy is measured by passing the input speech signal (for the frame in question) through an all-zero inverse filter representing the reference data. If the reference data match the input data, then the spectral notches in the inverse filter will match the spectral peaks in the input signal and a low energy residual output will result. This residual energy is normalized to a value greater than one, by dividing by the residual energy which results when the inverse filter is optimally matched to the input data.

The prediction residual is computed easily as the inner product of the autocorrelation function of the input with the autocorrelation function of the inverse filter, as will be shown later. Normalization by the residual of the input signal is not so simple. In essence the autocorrelation matrix must be inverted, and the traditional method of choice is Levinson's algorithm. An improvement on this algorithm which limits intermediate computations to a magnitude less than 1 was demonstrated by LeRoux and Geuguen.[16]

B. MODIFICATIONS TO THE INPUT AUTOCORRELATION FUNCTION

Initial experiments indicated the desirability of modifying the similarity measure. These modifications, however, were actually imple-

mented by adding some preprocessing to the input autocorrelation sequence.

First, an inherent problem in using inverse filtering techniques for comparing speech data arises when substantial modeling of very low amplitude parts of the spectrum takes place. Since an inverse filter can exhibit an extremely high gain at certain frequencies (especially at high frequencies), imperceptibly small noise signals in the input at these frequencies can strongly dominate the residual energy output from the inverse filter.

A solution to this problem is to add a "noise floor" to the speech signal to prevent very high gain in the inverse filter. This is done by adding uncorrelated white noise with energy proportional to the input signal energy. This is implemented by multipying the frame energy [r(0)] by some factor (1.1, in this case) and leaving all other autocorrelation terms unchanged. Multiplying by 1.1 corresponds to adding 10% noise to the signal and reduces the range in spectral amplitude to about 20 dB. This algorithm improves the performance of the verification system in a noisy environment as it reduces the effect of the dissimilarity between the lower energy portions of the speech which have been corrupted by background noise. The choice of the level for the noise floor represents a compromise between the system performance in low and high background noise environments. Figure 22 illustrates the effect of multiplying the zeroth-lag autocorrelation term [r(0)] by 1.1 for a nineteenth-order LPC modeled spectrum. The dashed curve is for r(0) increased by 10%, showing the reduction in dynamic range of the speech signal.

The second modification was to regress out gross energy trends versus frequency, similar to the method used in previous filter bank work that used Legendre polynomials and sine/cosine basis functions. The method used here, however, was to filter the signal with its own low-order (first-order) inverse filter.

The side-effects of these two operations are (1) the reduction of the dynamic range of the spectral amplitude to about 20 dB and (2) the need to use a higher-order LPC analysis to accurately model the formant locations. (Instead of our usual 14th-order model, a 21st-order model was used, reducing to 20th-order after regression.)

C. TIME REGISTRATION

In order to compute the similarity measure between the incoming speech and stored reference patterns, it is necessary to compensate for changes in the length and timing of the input utterance. Two methods of time registration were used during these experiments.

The first method was to compare "scanning" patterns formatted from the input data to reference scanning patterns. This method is identical to that used previously for filter bank parameters. These input scanning patterns for each word were formed by using the decision function for six frames, spaced 20 ms apart, at times of -50, -30, -10, +10, +30,

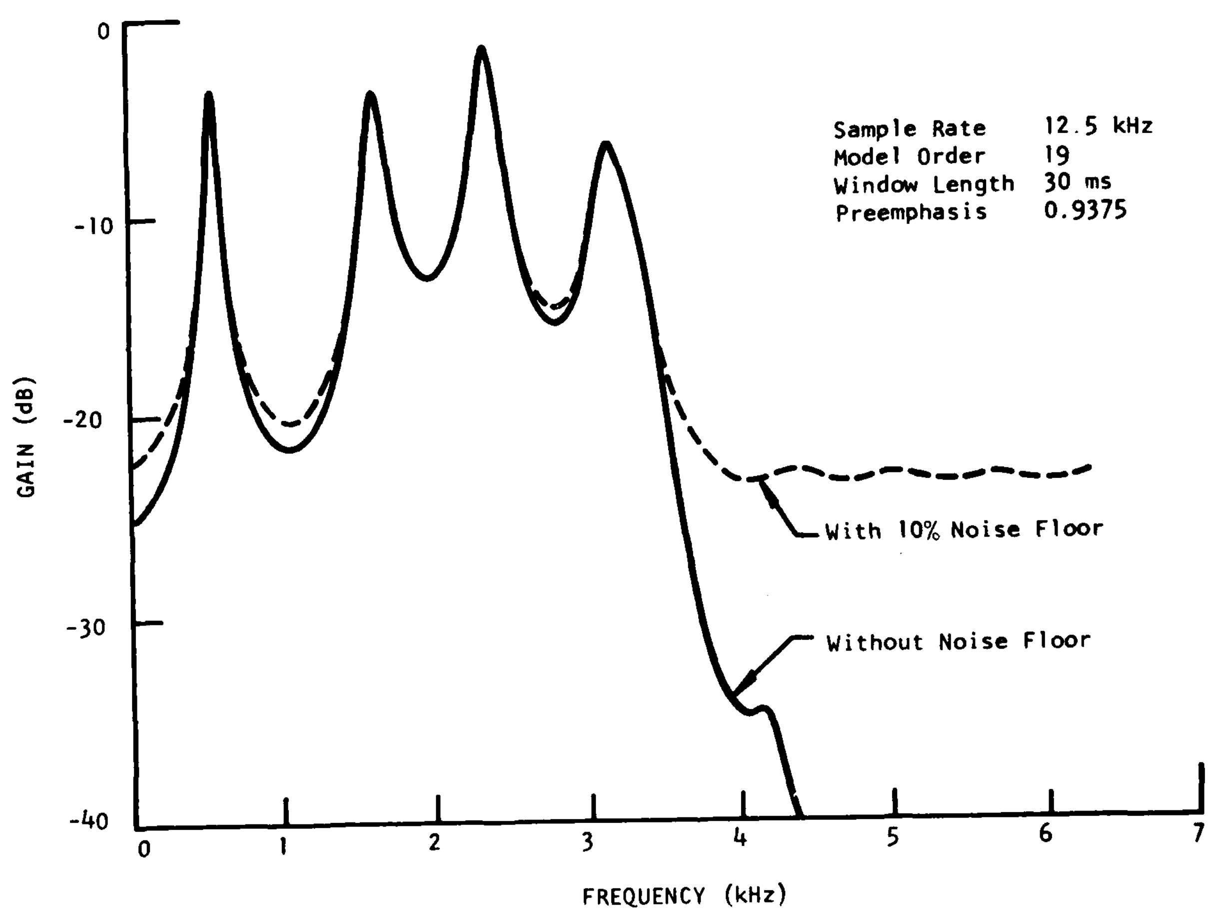


Figure 22 Comparison of LPC Modelled Spectra With and Without Noise Floor

and +50 ms around each input sample time. A scanning pattern centered about frame n thus consisted of the frame sequence (n-5, n-3, n-1, n+1, n+3, n+5). The reference patterns were formed at the same time intervals around the energy peak in each of the reference words. A scanning error was then computed between each input and reference scanning pattern. In other words in this method, there was no time normalization for each individual word; however, the times between words were nonlinearly scaled.

The second method used was a modification of the basic dynamic programming algorithm used by Itakura, which performs word specific time registration so as to maximize the similarity of input and reference data for each reference word in turn. There were three significant modifications to this technique as used in this experiment. First, and most significant, is that endpoints are unconstrained. That is, there is no algorithm that arbitrarily (or otherwise) constrains the dynamic optimization routine to start and end on specific input frames. Processing time is substantially increased by elimination of these constraints. However, the reliability burden on endpoint determination no longer exists.

A second difference is that the reference data are represented only at every other input frame. That is, the frame period of the reference data is twice the frame period of the input data. Our experience, at least for word recognition, suggests that performance is relatively insensitive to the number of reference frames per vocabulary word above a threshold level of about 20 frames per second. This frame skipping in the reference data does two things. First, it halves the amount of reference data that must be stored. Second, it halves the number of dynamic programming computations that must be performed.

The final difference is that penalty errors are added when nonlinear warping occurs. Figure 23 gives a graphic comparison of the basic and modified dynamic programming techniques.

D. DERIVATION OF THE RESIDUAL ERROR DECISION FUNCTION

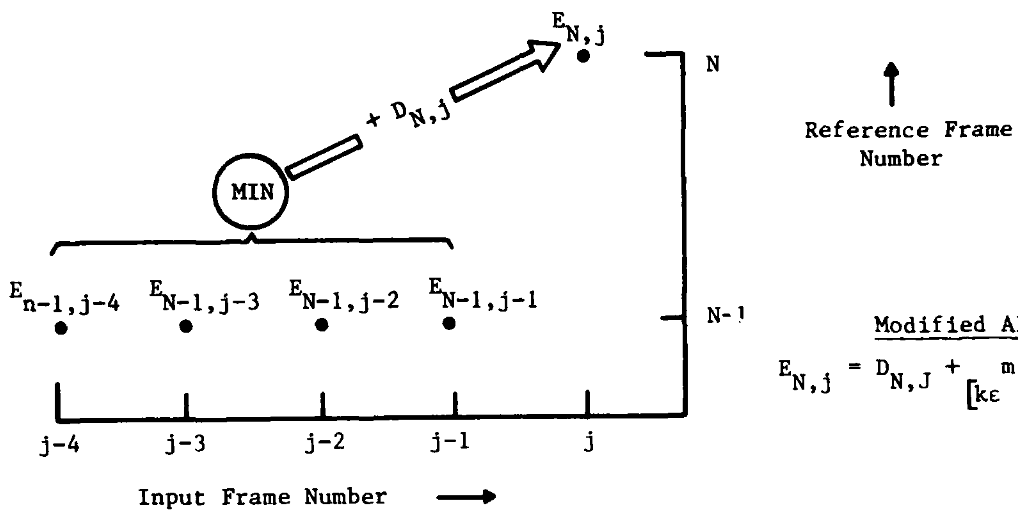
The frame-to-frame comparison used in this investigation is the energy of the residual signal left after passing an input through a linear predictive inverse filter, normalized by the energy in the input signal. The decision function then is the ratio of the normalized residual energy obtained when the unknown input is passed through the inverse filter for the reference, to the normalized residual energy obtained when the unknown input is passed through its own inverse filter.

In general, the autocorrelation (energy) of the output (y) of a filter can be found* from the autocorrelation of the original signal by

$$Ey = r(i) * h(i) * h(-i)$$

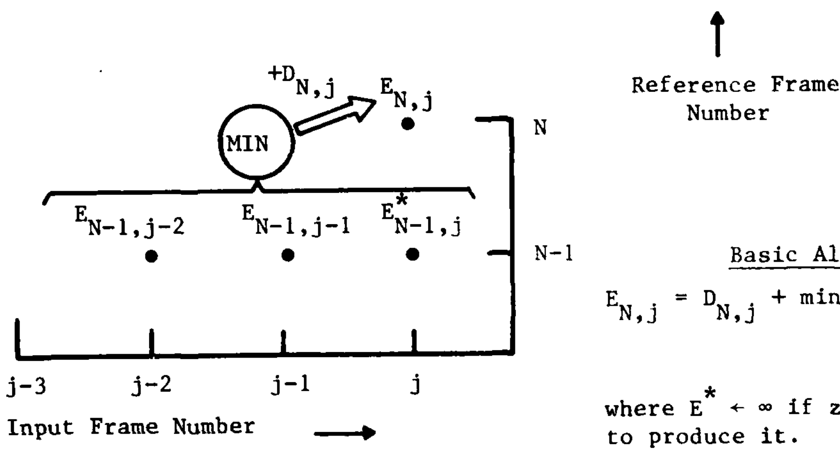
---

* Equation 10-36 of Athanasios Papoulis, Probability, Random Variables, and Stochastic Processes, Reference 17.



<u>Modified Algorithm</u>:

$$E_{N,j} = D_{N,J} + \min_{[k\varepsilon\ 1,4]} \left[W_k * E_{N-1,j-k}\right]$$



<u>Basic Algorithm</u>:

$$E_{N,j} = D_{N,j} + \min(E^*_{N-1,j}, E_{N-1,j-1}, E_{N-1,j-2})$$

where $E^* \leftarrow \infty$ if zero lag were used to produce it.

Figure 23 Comparison of Basic and Modified Dynamic Programming Algorithms

- $D_{N,j}$ is the distance measure between reference frame N and input frame j.
- $E_{N,j}$ is the optimum (minimum) subsequence distance up through reference frame N, given that input frame j corresponds to reference frame N.

where h(i) is the impulse response of the (inverse) filter, r(i) is the autocorrelation of the input [in our case, the input has been preemphasized [x(i)-x(i-1)] and windowed with a Hamming window], and "*" indicates convolution. However, since the convolution, h(i) * h(-i), is just the autocorrelation [(0)] of the impulse response, Ey is just the convolution of the autocorrelation of the input with the autocorrelation of the impulse response.

A more detailed derivation of the expressions for these residual energies can be given by considering passing a discrete-time input sequence $\{x(n)\}$ of length L, through an inverse filter of the form

$$A(z) = 1 + \sum_{i=1}^{M} a(i) z^{-1} = \sum_{i=0}^{M} a(i) z^{-1} \qquad \text{with } a(0) = 1.$$

The output, y(n), of this filter is x(n) + a(1)x(n-1) + ... + a(M)x(n-M). The total energy in the filtered signal, y, (i.e., the residual energy or the energy in the residual signal) is given by

$$Ey = \left[\sum_{n=0}^{L} \sum_{i=0}^{M} a(i)x(n-i) \right]^2 = \sum_{n=0}^{L} \left[\sum_{i=0}^{M} a(i)x(n-i) \sum_{k=0}^{M} a(k)x(n-k) \right],$$

or alternatively, as

$$Ey = \sum_{i=0}^{M} \sum_{k=0}^{M} a(i)a(k) \sum_{n=0}^{L} x(n-i)x(n-k) = \sum_{i=0}^{M} \sum_{k=0}^{M} a(i)a(k)r(i-k),$$

where r(i-k) is the autocorrelation of the input signal, x. When the input signal is the same as that used to derive the a(i)'s,*

$$\sum_{i=1}^{M} a(i)r(i-k) = - r(k) \qquad k = \{1,2,\ldots,M\},$$

---

* Equation 3 of J.D. Markel, "Digital Inverse Filtering - A New Tool for Formant Trajectory Estimation," Reference 18.

in which case, the residual energy becomes

$$Ey' = \sum_{i=0}^{M} a(i)r(i) \; .$$

Both Ey and Ey′ can then be normalized by dividing by r(0), the energy in the input signal, or equivalently, by letting r(i) be a normalized autocorrelation:

$$r(i) = \sum_{n=0}^{L-i} x(n)x(n+i) \Bigg/ \sum_{n=0}^{L} x^2(n) \; .$$

Note that Ey′ is just the inner product of two vectors, $< \vec{a}, \vec{r} >$. With some manipulation, Ey can be written as

$$Ey = 2 \sum_{i=0}^{M} \left[\sum_{j=0}^{M-i} a(j)a(j+i) \right] r(i) - \sum_{j=0}^{M} a^2(j)r(0) \; .$$

If $\vec{\rho}$ is the column vector $(\rho(0), \rho(1), \ldots \rho(M))$, where

$$\rho(i) = \sum_{j=0}^{M-i} a(j)a(j+i) \quad ,$$

then Ey can also be written in inner product form as $2<\vec{\rho}, \vec{r}> - \rho(0)r(0)$.

The (log of the) ratio of Ey to Ey′ was proposed by Itakura [15] for use as a decision function for isolated word recognition and was used by Wakita [19] for vowel recognition.

Wakita points out that Ey can represent the residual energy for either an arbitrary input passed through the inverse filter for a reference, or a reference passed through the inverse filter for the arbitrary input. Ey′ is the residual energy for any input passed through its own inverse filter. The decision function used in this investigation was a modification to the ratio of Ey (an arbitrary input passed through the inverse filter for the reference) to Ey′ (the arbitrary input passed through its own inverse filter).

E. THE EXPERIMENTAL DATA SET

The data set used during all of the early experiments was the same as that used in the filter bank experiments done for testing the changes made to the upgraded system. This set was collected prior to this contract and consisted of one enrollment session (48 phrases) and 21 verification sessions (12 phrases each) from 11 speakers (6 males, 5 females).

The collection of data from additional speakers to expand the number of speakers in the data set was completed during this contract. These data also consisted of one enrollment session and 21 verification sessions from 24 additional speakers (12 males, 12 females), including one pair of identical twins (males). With the eleven original speakers, there are data from 18 males and 17 females. This gives data for 12,138 (18*17*21 + 17*16*21) type II trials and 735 (35*21) true speaker trials. All the data were preprocessed to produce the LPC analysis files used in the residual error speaker verification experiments. However, due to the processing time required for this experiment, only two verification sessions from each speaker were processed for impostor trials. Also, in order to avoid data dependencies between the two twins, only one twin was used in the experiments. This yielded 13,056 (12*2*2*17*16) single-phrase type II trials.

F. FIXED PATTERN TIME NORMALIZATION EXPERIMENTS

The first scheme used for time-normalization was to compare "scanning" patterns formatted from the input data to reference scanning patterns. These input scanning patterns were formed by using the decision function at times of -50, -30, -10, +10, +30, and +50 ms around each input sample time. The reference patterns were formed at the same time intervals around the energy peak in each of the reference words. A scanning error was then computed for each input time sample, "j", as

$$SE(j) = \left[1/6 \sum_{k=1}^{6} \frac{Ey(j+2k-7)}{Ey'(j+2k-7)} - 1 \right] * 1000 ,$$

where both Ey and Ey′ were calculated using the modified input autocorrelations. Ey was calculated directly as explained earlier; however, Ey′ was a byproduct of a general subroutine that iteratively calculated the predictor coefficients. Note that since the usual range of Ey/Ey′ is 1.0 through 1.4, (Ey/Ey′)-1 is fairly close to the log(Ey/Ey′) used by Itakura. The effect of not taking the log is to penalize the larger deviations more severely.

A complete speaker verification experiment was run on the 11-speaker, 21-session data set using scanning errors (SE(j)′s). The minima of the scanning errors for all words in a phrase were combined to find the sequence of scanning error minima having the smallest total se-

quence error according to the following equation:

$$\text{OPTSEQ ERROR} = \sum_{i=1}^{3} (1 + \alpha\ m(i))(1 + \alpha\ m(i+1)) \left[1 + \beta \left[\frac{dt(i) - dthat(i)}{dthat(i)} \right]^2 \right]$$

where m(i) is a (local) minima of SE(j) for the word in position i,
dt(i) = t(i+1) - t(i) , (t(i+1) must be > t(i)),
dthat(i) is the expected value of dt(i) for the speaker,
α = 0.01, and
β = 1.0.

Plots of the average (e) of the m(i)'s in the sequence having the lowest OPTSEQ ERROR for both true speakers and impostors for the residual energy measure are shown in Figure 24 for males and Figure 25 for females. Also shown in Figures 24 and 25 are the comparable curves using SE(j)'s derived from Euclidian distances between input and reference spectral patterns.

Plots of the normalized errors (e/ehat) of the m(i)'s for the residual energy measure are shown in Figure 26 for males and Figure 27 for females. Again, the curves derived from spectral SE(j)'s are also shown in both figures.

There are several important characteristics to note from these figures:

1. Performance using the residual energy measure is superior to that for the spectral measure.

2. Performance for males is superior to that for females for the residual energy measure for both e and e/ehat, and for the spectral measure for e.

3. Performance using normalized errors is better for the spectral measure and is worse for the residual energy measure.

4. The number of unregistered phrases was lower for both true speakers and impostors using the residual energy measure rather than the spectral measure.

This experiment was next rerun on the 11-speaker data base using thresholds on the scanning error of 500 and on the OPTSEQ error of 50. The prior experiment had no upper limits on either of these parameters. The exciting result was the drastic increase in the number of phrases not registered for impostors, with very little sacrifice in the number of phrases not registered for true speakers, as shown in Table 36. Also given in Table 1 are the equal error rates for each case, with unregistered phrases for true speakers not counted.

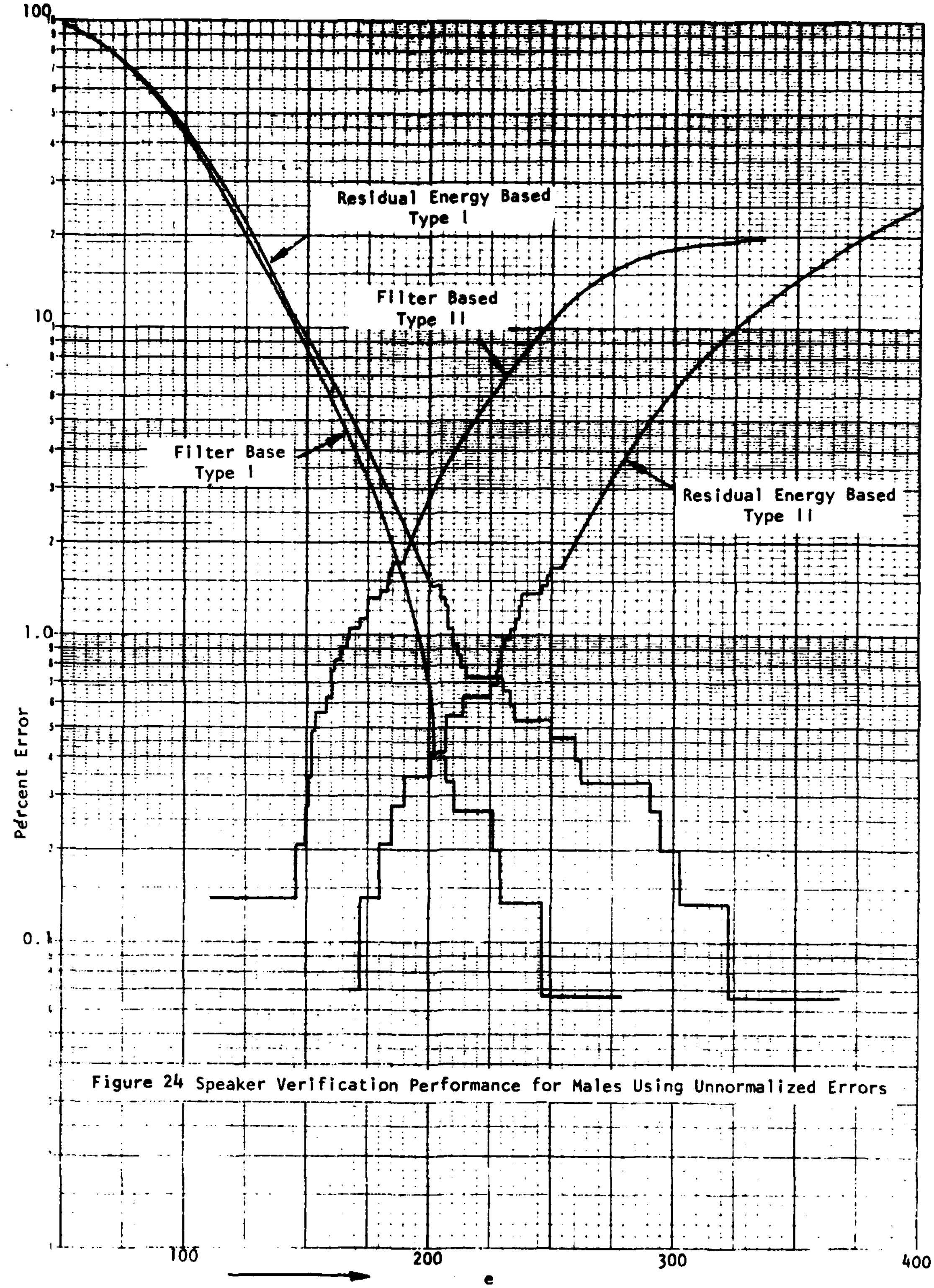


Figure 24 Speaker Verification Performance for Males Using Unnormalized Errors

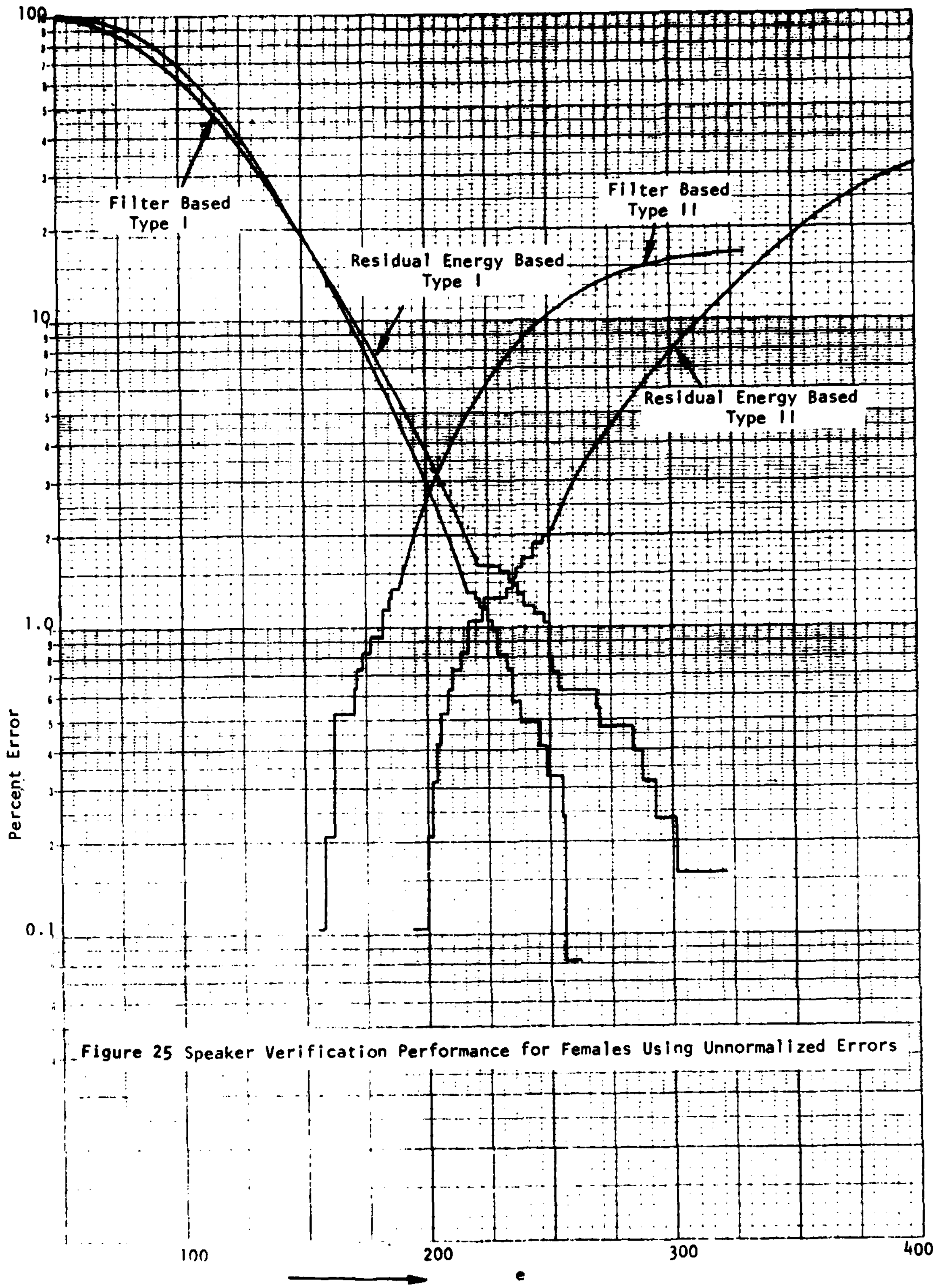


Figure 25 Speaker Verification Performance for Females Using Unnormalized Errors

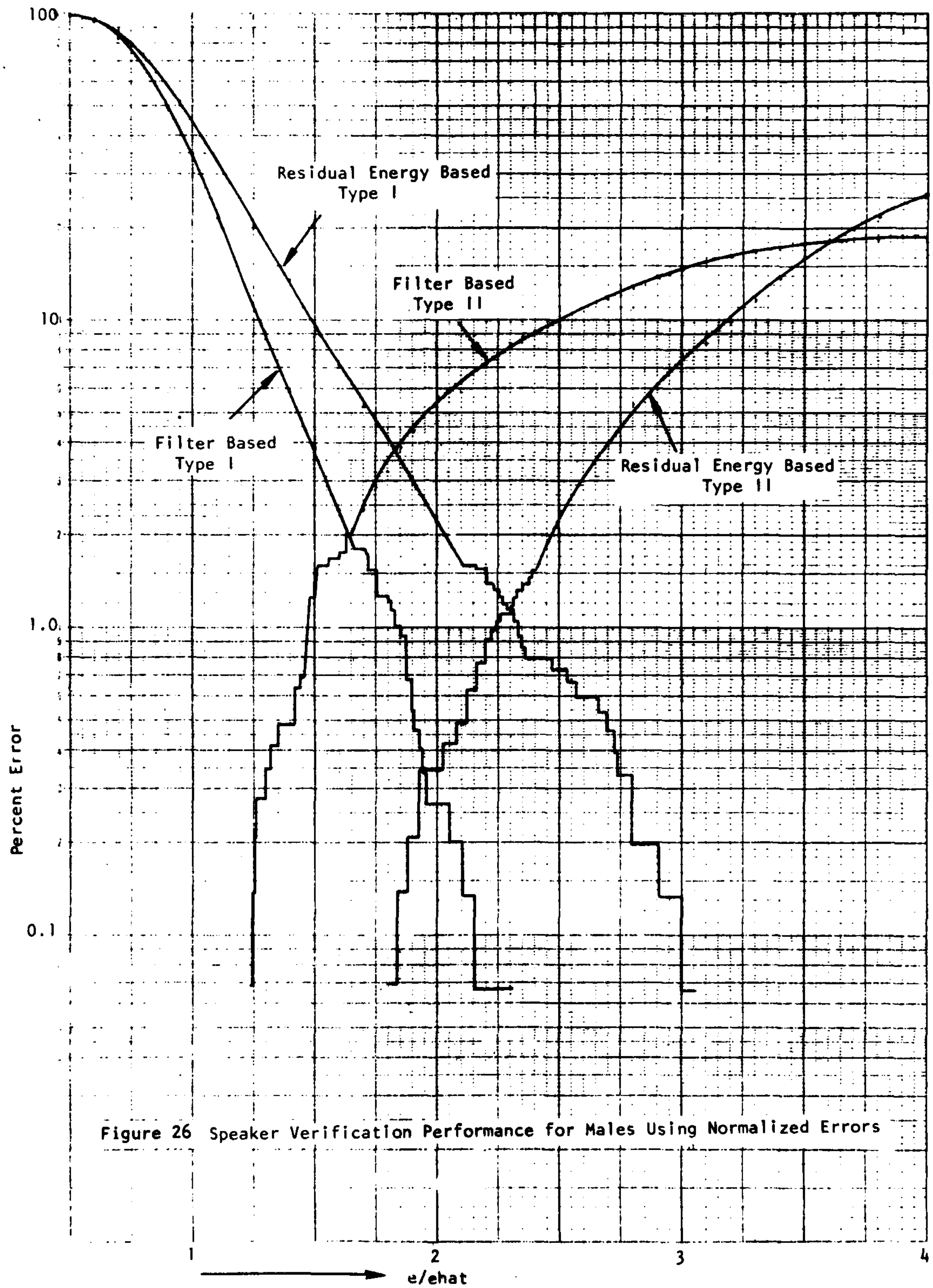


Figure 26 Speaker Verification Performance for Males Using Normalized Errors

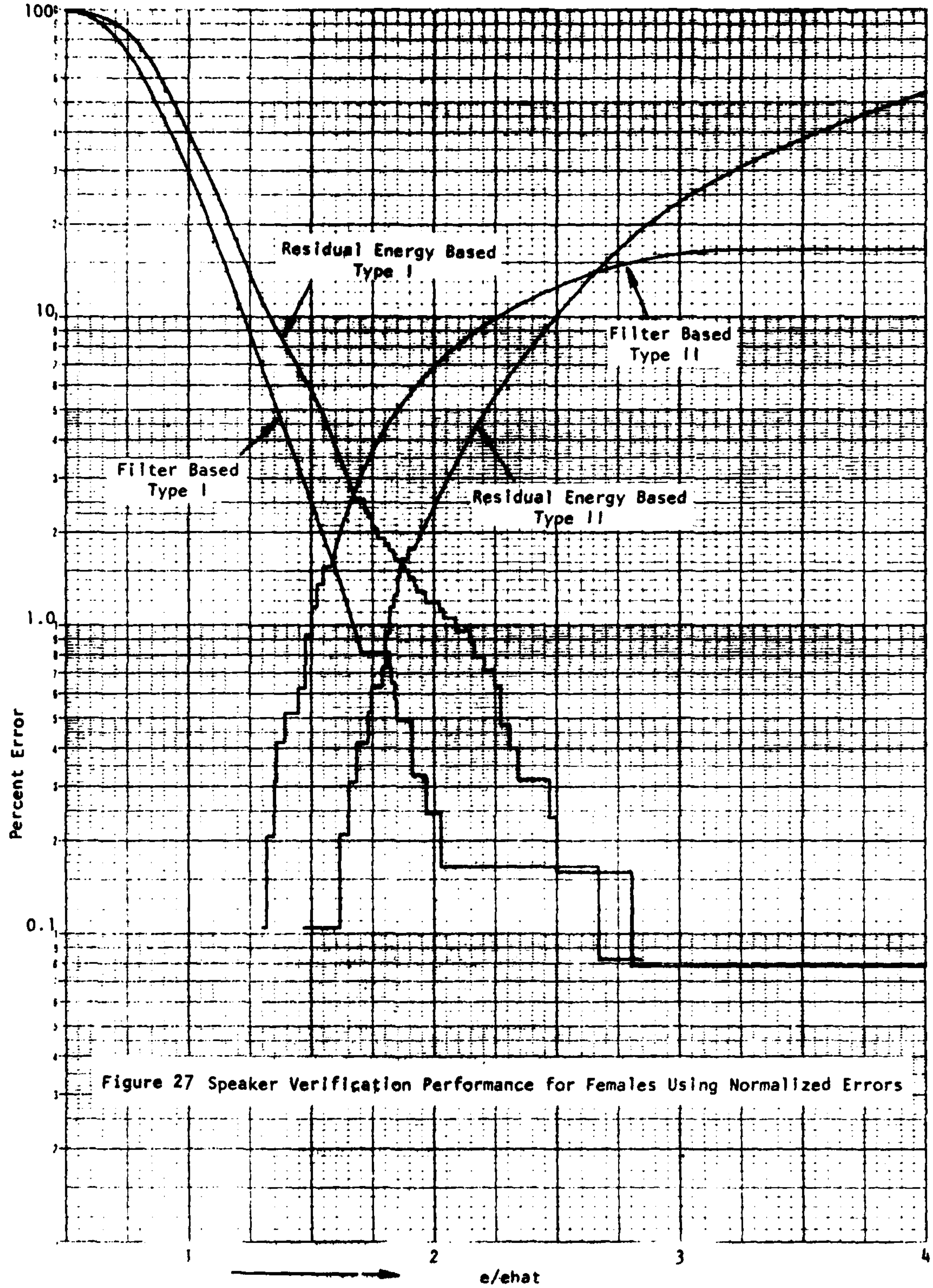


Figure 27 Speaker Verification Performance for Females Using Normalized Errors

TABLE 36. UNREGISTERED PHRASE RATES FOR IMPOSTORS/TRUE SPEAKERS AND EQUAL ERROR RATES IN PERCENT

FEATURE TYPE:	Filter bank	Residual Error	Residual Error
THRESHOLDS? :	yes	no	yes
% UNREGISTERED PHRASES (IM/TS):			
Males :	80.8/1.0	4.4/0.1	93.8/0.3
Females :	83.2/2.7	1.7/0.0	92.6/0.3
Total :	81.8/1.8	3.3/0.0	93.3/0.3
EERs IN % USING UNNORMALIZED ERRORS:			
Males :	1.74	0.73	0.60
Females :	3.00	1.35	1.27
Total :	2.20	1.08	0.83
EERs IN % USING NORMALIZED ERRORS:			
Males :	2.01	1.12	0.93
Females :	1.63	1.59	1.35
Total :	1.79	1.55	1.33

Next, two additional experiments with the eleven-speaker data set were done. One experiment reversed the order of regressing the input and applying the noise floor. The second reversed the roles of the input and the reference, resulting in a decision function that was the ratio of the residual error from the reference passed through the optimal inverse filter for the input, to the residual error from the reference passed through the optimal inverse filter for itself.

Table 37 compares the results of these two residual error experiments with the results of the prior experiment (middle column of Table 36). Note that both of these experiments yielded poorer results. Both experiments were done without thresholds applied. Figure 28 shows the more descriptive Type I/Type II curves for these experiments.

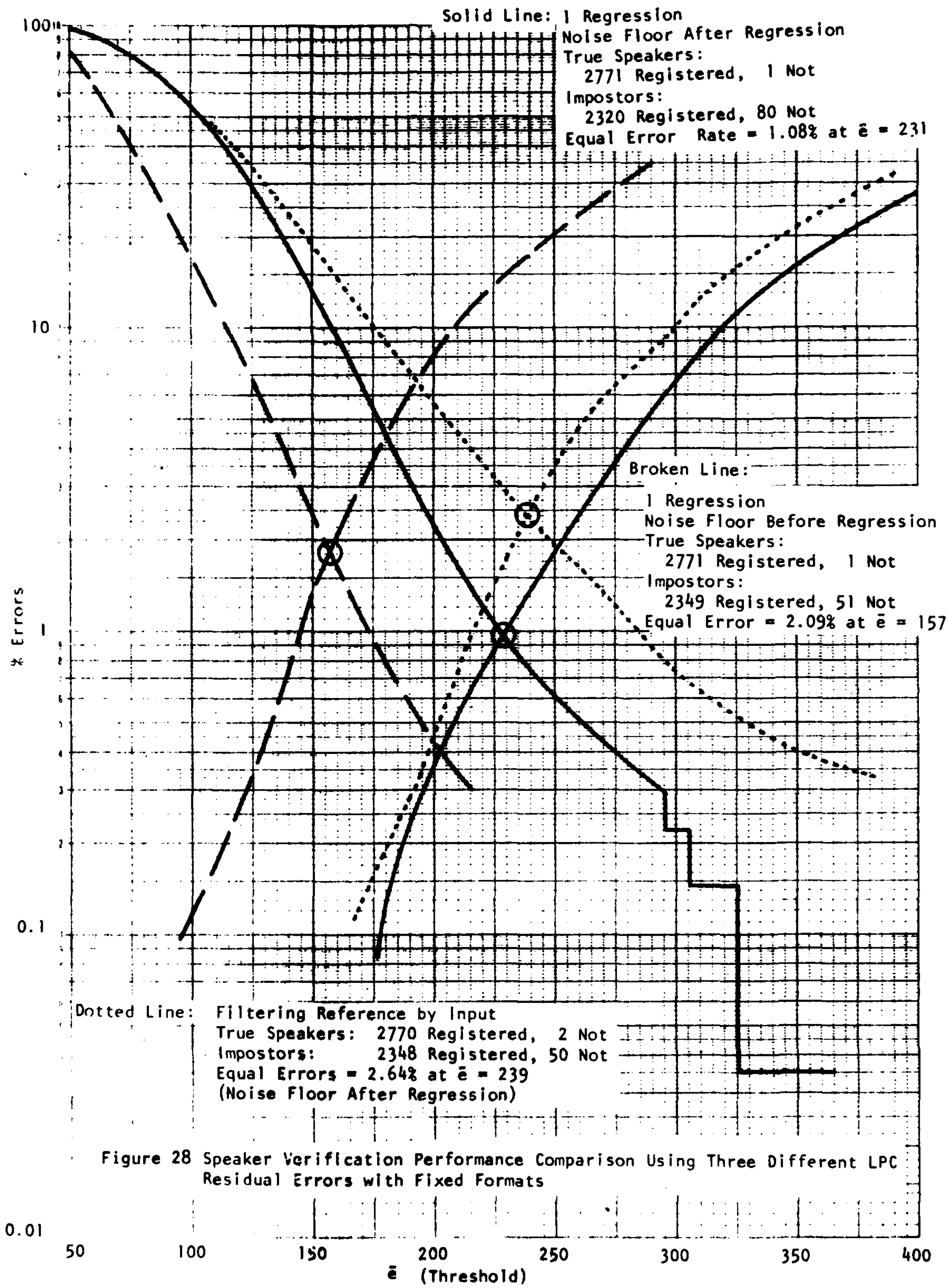


Figure 28 Speaker Verification Performance Comparison Using Three Different LPC Residual Errors with Fixed Formats

TABLE 37. UNREGISTERED PHRASE RATES FOR IMPOSTORS/TRUE SPEAKERS AND EQUAL ERROR RATES IN PERCENT

X FILTERED BY Y:	Input by ref	Ref by input	Input by ref
REGRESS BEFORE /AFTER NOISE FL:	before	before	after
THRESHOLDS? :	no	no	no
% UNREGISTERED PHRASES (IM/TS):			
Males :	4.4/0.1	2.9/0.1	3.1/0.1
Females :	1.7/0.0	0.7/0.0	0.6/0.0
Total :	3.3/0.0	2.1/0.1	2.1/0.0
EER's IN % USING UNNORMALIZED ERRORS:			
Males :	0.73	2.71	1.11
Females :	1.35	2.62	2.50
Total :	1.08	2.64	2.09
EER's IN % USING NORMALIZED ERRORS:			
Males :	1.12	2.36	1.81
Females :	1.59	3.18	1.35
Total :	1.55	3.08	1.91

F. NONLINEAR TIME WARPING EXPERIMENTS

Of even greater interest are the results of another set of experiments. Instead of using reference patterns formatted at fixed intervals around the four energy peaks in an input, a nonlinear time warping with dynamic programming was used for time alignment of the input to the reference. As expected, because of better time alignment and larger reference patterns, there was indeed another dramatic improvement in performance. Figures 29 and 30 compare the performance on the 11-speaker data set using (1) the DSG filter bank and 6-column fixed format reference patterns, (2) the residual error based, 6-column fixed format reference patterns, and (3) the residual error based reference patterns with nonlinear time warping of the input. Although not shown, applying a threshold to the nonlinear time warped case shifts the bottom of the true speaker curve to the left and flattens the top of the impostor curve, since only 1.8% of the impostor phrases even "register," i.e. for only 1.8% of the phrases, the match between the input and the reference is less than a given threshold. It should be noted that the nonlinear time warping method does require about four times as much memory for storing reference patterns as the other two methods.

All three of these curves are for decisions based only on one phrase. Two experiments were run on the 11-speaker data set with a two-phrase strategy using all pairs of the twelve phrases from each im-

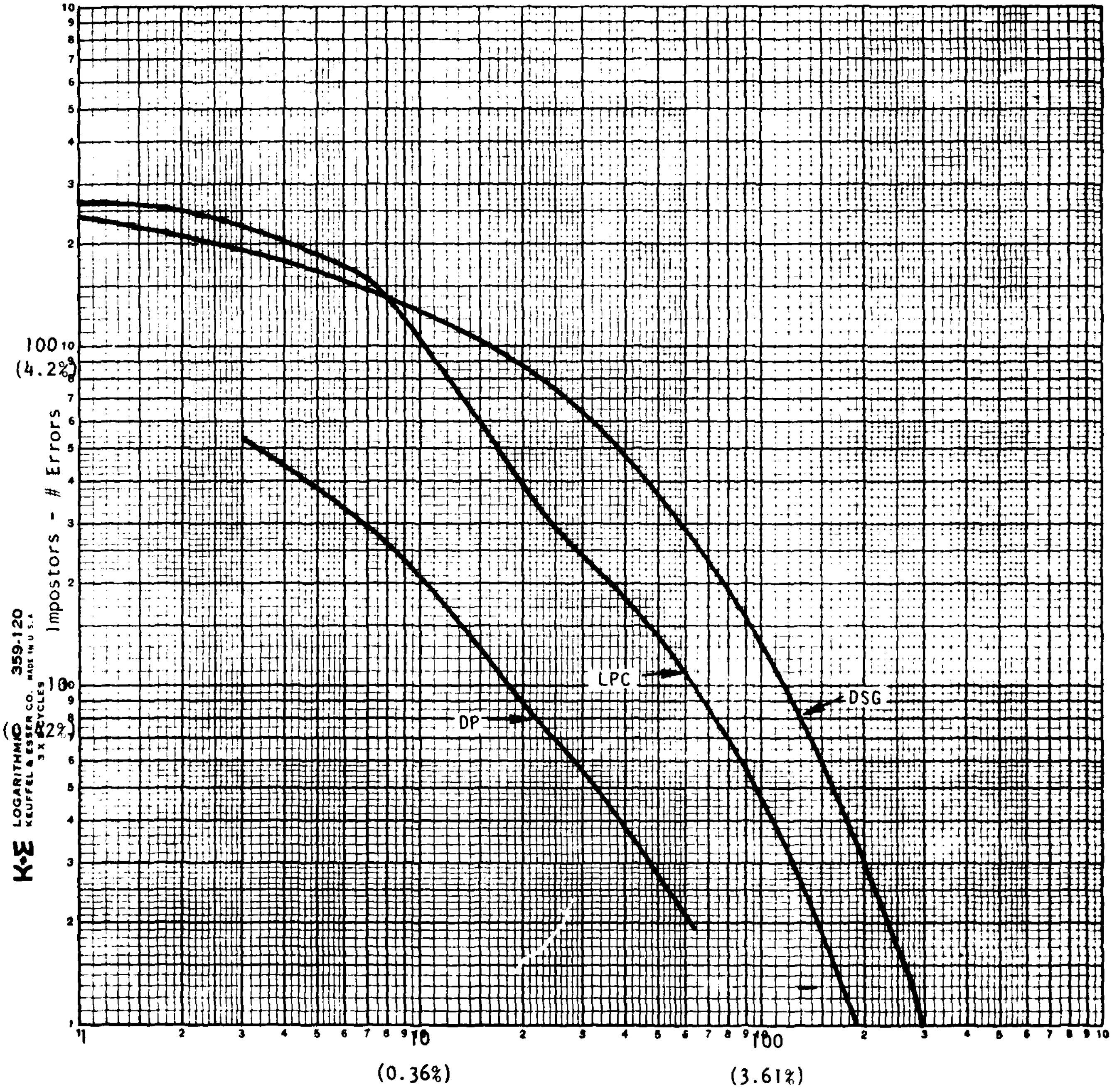


Figure 29 Speaker Verification Performance Comparison Using (1) Filter Bank, (2) LPC Residual Errors, and (3) LPC Residual Errors with Nonlinear Time Warping

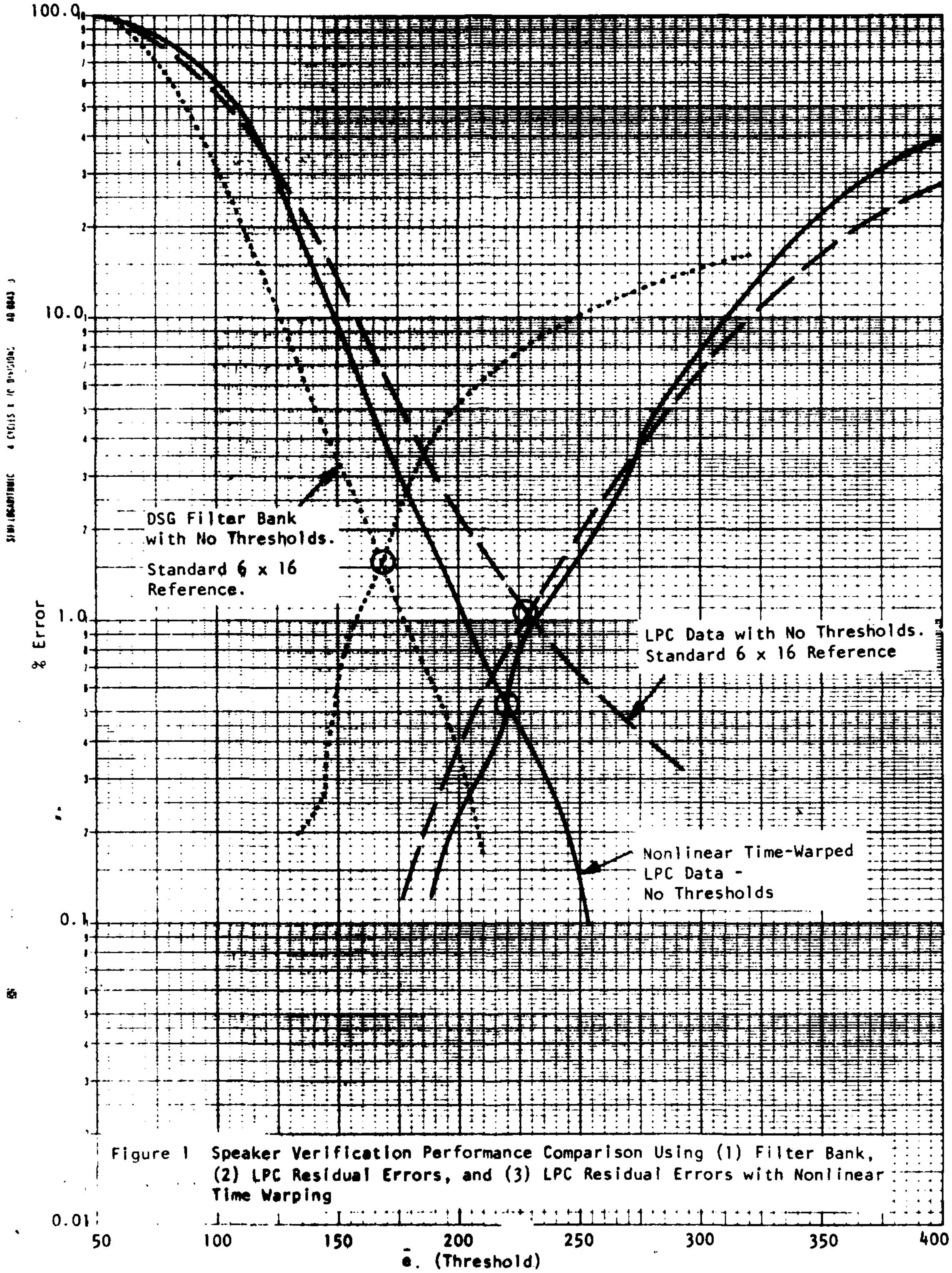


Figure 1 Speaker Verification Performance Comparison Using (1) Filter Bank, (2) LPC Residual Errors, and (3) LPC Residual Errors with Nonlinear Time Warping

postor and true speaker session. The equal error rate was less than 0.1%; however, since this represents only one or two errors, it is not possible to estimate the actual equal error rate with a high degree of confidence. This emphasized the need to expand the test data set from eleven speakers. The preliminary results, however, suggested that a nonlinearly time warped residual error technique can achieve the goals of <1.0% true speaker error and <0.1% impostor error called for in the statement of work.

In the method for performing the nonlinear time warping in this experiment, boundary points were manually chosen between each of the four words in the first four enrollment phrases (one sample of each of the 16 words), based primarily on energy contours. A reference pattern for each word was then formatted by excising the LPC parameters every other time sample.

The procedure used on the remaining sixteen phrases during enrollment was the same as used during verification. One "super" reference pattern was formatted by concatenating the reference patterns for the four words. This super reference pattern was then used in a dynamic time-warping routine to find the best match between the input and the reference. In this procedure, the Itakura type distance described earlier was calculated between each input frame and each reference frame. Running cumulative distances were calculated for every input and reference frame, and the minimum cumulative error across all input frames for the last reference frame was selected and the minimum error path was backtracked. One constraint on the calculation of the cumulative distances was that no two input frames were allowed in any path for the same reference frame, nor could the input frames corresponding to adjacent reference frames be farther apart than four 10 ms frames. Also, the increment to the cumulative error between reference frames was weighted by the distance between input frames. This weighting was determined from 0.2 times the square of the natural log of half the difference in the input frame numbers. This results in the following weights:

input frame distance:	1	2	3	4
weight:	0.096	0.0	0.033	0.096

The warped input patterns for each of the four words were then averaged in with the reference pattern in a 1/2, 1/3, 1/4, and 1/5 proportion, depending on whether the input was from the 2nd, 3rd, 4th, or 5th repetition of the word. The number of time samples in the reference pattern did not change after the initial patterns had been established. Since no thresholds were used during enrollment, every phrase was registered.

During verification, an extra calculation was added, since the cumulative error in the dynamically time-warped path was not the score used in the decision. Instead, the score from each selected frame of the input was weighted by the clipped energy of the frame. This gave less weight to the lower energy frames, while the clipping prevented the

high energy vowels from totally dominating the decision function. Hence, the actual decision function was given by

$$D = \frac{\sum_{k} d(k)\ [\min\ (e(k)\ ,\ 2\ \langle e\rangle)]}{\sum_{k} \min\ [e(k)\ ,\ 2\ \langle e\rangle]} ,$$

where d(k) was the difference between reference frame k and the selected frame of the input data; e(k) was the energy in the preemphasized, windowed, input frame (the square root of $\rho(0)$ divided by the number of samples in the input frame) corresponding to frame k; and <e> was the average of all of the e(k) over the input utterance.

G. AUTOMATIC ENROLLMENT

Programs were next developed to do automatic (no manual intervention) enrollment with residual error data, and an experiment using the 11-speaker data set was run to determine type I/type II performance. The pleasing result of this experiment was that there was no degradation in performance using automatic enrollment.

The change made in this experiment was to automate the selection of boundary points between words, a task that previously had been done manually. The rest of the algorithm remained unchanged. In this automatic enrollment, the four monosyllabic words were first located by finding the four largest peaks in a "smoothed" energy function, defined as the sum of a function of the RMS energies from six (10 ms) frames before the current frame through six frames after the current frame. In the filter bank approach, the RMS energy used was a function only of filters 3 through 12, which eliminated those portions of the spectrum having high energies during either nasals (the low frequencies) or sibilants (the high frequencies). Since in the LPC analysis being used in these experiments, we have only autocorrelations of the inputs, another method must be found to eliminate the contributions to the energy from both the very low and the very high frequencies. This could have been accomplished by passing the original signal through a bandpass filter before calculating the autocorrelations. Alternatively, the autocorrelation (r(i)) of such a bandpassed signal can be found from the autocorrelation of the original signal by the same technique used to calculate the residual energy out of the inverse filter as explained earlier. This can be done since

$$r'(i) = r(i) * h(i) * h(-i),$$

where h(i) is the impulse response of the bandpass filter, r(i) is the autocorrelation of the input [in our case, the input has been preemphasized [x(i)-x(i-1)] and windowed with a Hamming window], and "*" indicates convolution. Remember, however, that since the convolution, h(i)

* h(-i), is just the autocorrelation [ρ (0)] of the impulse response, $r'(i)$ is just the convolution of the autocorrelation of the input with the autocorrelation of the impulse response. A simple filter to eliminate the energies at the low and high ends would be one with a zero at 0 Hz and two zeros at 5000 Hz. The transfer function for such a filter is

$$H(z) = [1 - 2\cos(a) z^{-1} + z^{-2}] (1 - z^{-1}),$$

where "a" is two pi times 5000 Hz divided by the sample frequency (12500 Hz), yielding 2cos(a) = -1.618034. This corresponds to an impulse response, h(t), of

$$h(t) = del(t) - [1+2\cos(a)]\ del(t-1) + [1+2\cos(a)]\ del(t-2)$$

$$- del(t-3)$$

$$= del(t) - 0.618034\ del(t-1) + 0.618034\ del(t-2) - del(t-3),$$

where "del" is the delta function. h(t) * h(-t) then becomes

$$h(t)*h(-t) = - del(t-3) - 1.236068\ del(t-2) + 0.854102\ del(t-1)$$

$$+ 2.763932\ del(t) + 0.854102\ del(t+1)$$

$$- 1.236068\ del(t+2) - del(t+3).$$

Substituting this into the equation for $r(i)'$, with i = 0, gives

$$r(0)' = 2.763932\ r(0) + 2\ [0.854120\ r(1) - 1.236068\ r(2) - r(3)]$$

$$= 2.763932\ r(0) + 1.7082\ r(1) - 2.4721\ r(2) - 2\ r(3).$$

As a final step, $r(0)'$ is weighted by (1 + r(1)/r(0)), a function which varies linearly with the mean frequency of the input signal from two at 0 Hz to zero at the half sample frequency (6250 Hz). The "smoothed" energy actually used is then the square root of 0.01 times this value.

This "smoothed" energy is input to a peak and valley finding program that finds all peaks in the energy that have a value that is at least 500 and that is greater than 1.3 times the value of the preceding valley. The four largest peaks (independent of time differences between peaks) are retained as locations of the vowels in each of the four words. The beginning and end of the utterance for purposes of subsequent processing are determined as the times before the first energy peak and after the last energy peak, where the energy first drops below 25% of the peak value across the entire utterance. The boundaries between words are chosen as the valleys between the four largest energy peaks. Starting at the first 10 ms frame after the beginning of the utterance, reference patterns for each of the four words are built by extracting every other frame from the input and assigning it to the word defined by the valley point times. This is done for each of the first four phrases, which defines a reference pattern for each of the sixteen

words.

One step has been added to this procedure to refine the word boundary locations in the third and fourth phrases. The word order for the first four phrases insures that all possible phoneme transitions between words will have occurred in the first two phrases. Hence, a "mini-pattern" is formatted from the last three samples of the reference pattern for one word and the first three samples of the reference pattern for the following word, and is used to scan plus and minus twenty 10 ms frames around the valleys found in the input for the third and fourth phrases to refine the selected boundary locations between words and make them more consistent with those selected from the first two phrases. The selected boundaries between words in the third and fourth phrases are then moved from the valley points in the energy to the times of minimum scanning error found using the mini-patterns. Using these newly defined boundaries, reference patterns for these four new words are then selected as in the first two phrases.

The processing for all enrollment phrases after the fourth phrase and for all verification trials remained the same as described above for the case of manual enrollment.

H. EXPANDED DATA SET EXPERIMENTS

Next, to increase the confidence level of the results obtained in the prior experiments, the expanded, 34-speaker (17 males, 17 females, with only one twin used) data set was run using the automatic enrollment, nonlinearly time warped method. As expected, the performance on the larger data set was not as good as for the 11-speaker set, as can be seen from Table 38. The last two columns of Table 38 are for the expanded data set, with the last column showing the performance using a threshold large enough to allow almost all phrases to register. Note that since neither twin's data were consistent, the true speaker rejection rate was very high for both of the twins. The last column of Table 38 shows the equal error rates when both twins were excluded, as well as when only one twin was excluded.

TABLE 38. EQUAL ERROR RATES (EERs) IN PERCENT AND UNREGISTERED PHRASE RATES FOR IMPOSTORS/TRUE SPEAKERS FOR A ONE-PHRASE DECISION USING A NONLINEARLY TIME WARPED, RESIDUAL ENERGY FEATURE VECTOR

ENROLLMENT :	**MANUAL**	**MANUAL**	**AUTOMATIC**	**AUTOMATIC**	**AUTOMATIC**
NO. OF SPEAKERS:	11	11	11	34	34
THRESHOLD :	9E9	1.25	1.25	1.25	2.00
EERs IN %:					
Males :	0.595	0.556	0.486	0.804	1.517 (1.090)*
Females :	0.417	0.318	0.318	0.422	0.628
Total :	0.541	0.470	0.375	0.628	1.078 (0.836)*
EER THRESHOLDS:					
Males :	209	212	210	229	245 (233)*
Females :	240	237	233	226	230
Total :	221	221	216	228	240 (232)*
% UNREGISTERED PHRASES (IM/TS):					
Males :	0.0/0.0	97.9/0.1	98.1/0.1	98.3/1.3	3.2/0.0
Females :	0.0/0.0	99.4/0.3	99.4/0.4	98.7/0.4	1.2/0.0
Total :	0.0/0.0	98.5/0.2	98.6/0.3	98.5/0.8	2.2/0.0

* Values when both twins excluded. Unstarred values include only one of the two twins in the 35-speaker data set.

Additional experiments were run on this expanded data set to determine the effects of regression and model order. These results are shown in Table 39. Note that females had roughly three times the error rate when the regression was eliminated with relatively minor effect for the males. In contrast, although the error rate did increase when the model order was reduced, the effect was not nearly so great as when the regression was eliminated.

TABLE 39. EQUAL ERROR RATES IN PERCENT FOR ONE-PHRASE DECISION USING NONLINEARLY TIME WARPED, RESIDUAL ENERGY FEATURE VECTOR FOR VARIOUS REGRESSION/LPC ORDER COMBINATIONS

ENROLLMENT :	AUTOMATIC	AUTOMATIC	AUTOMATIC	AUTOMATIC
NO. OF SPEAKERS:	34	34	34	34
THRESHOLD :	2.0	2.0	2.0	2.0
LPC ORDER** :	21	20	15	14
REGRESS ORDER :	1	0	1	0
EERs IN %:				
Males :	1.517	1.532	1.746	2.148
	(1.090)*	(1.364)*	(1.351)*	(1.885)*
Females :	0.628	1.564	0.654	1.808
Total :	1.078	2.089	1.214	2.358
	(0.836)*	(1.941)*	(1.010)*	(2.297)*
EER THRESHOLDS:				
Males :	245	196	211	170
	(233)*	(193)*	(204)*	(168)*
Females :	230	160	201	139
Total :	240	175	208	153
	(232)*	(193)*	(203)*	(168)*

* Values when both twins excluded. Unstarred values include only one of the two twins in the 35-speaker data set.
** LPC order before regression.

The final experiment performed with a nonsequential decision strategy using the expanded data set was to test the results for a multi-phrase (but nonsequential) decision. These results are shown in Table 40 for the 34-speaker data set (i.e., one twin excluded) for both the tight (1.25) and the loose (2.0) thresholds.

TABLE 40. EQUAL ERROR RATES IN PERCENT AS A FUNCTION OF NUMBER OF PHRASES FOR NONLINEARLY TIME WARPED, RESIDUAL ENERGY FEATURE VECTORS

ENROLLMENT :	AUTOMATIC				AUTOMATIC			
NO. OF SPEAKERS:	34				34			
THRESHOLD :	1.25				2.0			
LPC ORDER* :	21				21			
REGRESS ORDER :	1				1			
NO. OF PHRASES:	1	2	3	4	1	2	3	4
EERs IN %:								
Males :	0.804	0.169	0.077	0.031	1.517	0.797	0.613	0.475
Females :	0.422	0.141	0.094	0.031	0.628	0.322	0.257	0.153
Total :	0.628	0.191	0.095	0.024	1.078	0.572	0.421	0.397
EER THRESHOLDS:								
Males :	229	220	216	212	245	250	253	260
Females :	226	213	211	210	230	234	233	225
Total :	228	218	213	211	240	242	243	241

* LPC order before regression.

I. SEQUENTIAL DECISION STRATEGY EXPERIMENTS

Finally, the 34-speaker data set was run using the sequential decision strategy employed in the current CIC system and in the VVU system. Of course, since these experiments used different data representations and preprocessings, the actual parameters used in the decision were different; however, the global strategy was identical. The phrase acceptance threshold for "registering" a phrase was 2.0 (the same as that used in the prior section), which resulted in all phrases being registered for both true speakers and impostors.

The results of these full decision strategy experiments are shown in Table 41 using three different sets of decision thresholds and allowing a maximum of either one or two repeats of any particular phrase. The number of true speaker trials was 714 (34 speakers * 21 sessions). However, in order to average out some of the phrase order dependence, the choice of the initial phrase was rotated around the 12 phrases, with the next eleven phrases following circularly after the initial phrase. Hence, there were actually 8,568 true speaker trials (714 * 12). The number of impostor trials was 1088 (17 impostors * 16 references * 2 sessions * 2 sexes). The same rotating of the choice for the initial phrase was done for the impostor data, resulting in 13,056 total impostor trials (1088 * 12). Note that all of these results are much better than even the desired goals given in the contract statement of work (1% true speaker rejection and 0.1% impostor acceptance).

TABLE 41. TRUE SPEAKER (TS) REJECTS AND IMPOSTOR (IM) ACCEPTANCES FOR A FULL "CIC," MULTIPHRASE DECISION STRATEGY USING A NONLINEARLY TIME WARPED, RESIDUAL ENERGY DECISION FUNCTION

TEST CONDITIONS:					
ENROLLMENT :	AUTOMATIC	AUTOMATIC	AUTOMATIC	AUTOMATIC	AUTOMATIC
NO. OF SPEAKERS:	34	34	34	34	34
NO. TS TRIALS :	8568	8568	8568	8568	8568
NO. IM TRIALS :	13056	13056	13056	13056	13056
PHR. ACCEPT. THRESHOLD :	2.00	2.00	2.00	2.00	2.00
MAX # PHRASES :	7	7	7	7	7
MAX # REPEATS :	1	2	1	2	1
VERIFICATION DECISION THRESHOLDS:					
	SET 1	SET 1	SET 2	SET 2	SET 3
NORMAL MODE:					
1 PHRASE :	1.15	1.15	1.15	1.15	1.15
2 PHRASES:	1.165	1.165	1.17	1.17	1.17
3 PHRASES:	1.18	1.18	1.19	1.19	1.19
4 PHRASES:	1.195	1.195	1.21	1.21	1.20
PREJUDICED MODE:					
1 PHRASE :	1.135	1.135	1.12	1.12	1.14
2 PHRASES:	1.155	1.155	1.15	1.15	1.16
3 PHRASES:	1.175	1.175	1.18	1.18	1.18
4 PHRASES:	1.195	1.195	1.21	1.21	1.20
TEST RESULTS:					
NO. OF TS REJECTS:					
Males :	47	37	30	28	38
Females :	12	15	3	6	11
Total :	59	52	33	34	49
NO. OF IM ACCEPTANCES:					
Males :	3	8	9	10	5
Females :	0	0	3	2	0
Total :	3	8	12	12	5
TOTAL (MALES & FEMALES) ERROR RATES IN %:					
TS Rejects:	0.689	0.607	0.385	0.397	0.572
IM Accepts:	0.023	0.061	0.092	0.092	0.038

* Only one of the twins in the 35-speaker data set was included; however, when impostor trials were run between twins, there were no successful impersonations.

REFERENCES

1. A.E. Rosenburg, "Automatic Speaker Verification: a Review," Proceedings of the IEEE, Vol. 64, No. 4, April 1976, pp. 475-487.

2. B.S. Atal, "Automatic Recognition of Speakers from Their Voices," Proceedings of the IEEE, Vol. 64, No. 4, April 1976, pp. 406-475.

3. G.R. Doddington, "Speaker Verification," RADC-TR-74-179, April 1974, 785135/5GI.

4. G.R. Doddington and B.M. Hydrick, "Speaker Verification II," RADC-TR-75-274, September 1975, A018901.

5. "Summary: Phase I - Analysis and Testing: Automatic Speaker Verification System," Contract No. F19628-74-C-0030, 30 July 1974.

6. A. Fejfar, "Test Results - Advanced Development Models of BISS Identity Verification Equipment; Volume I, Executive Summary," Mitre Technical Report, Mitre Technical Report, MTR-3442, Vol. 1, Rev. 1, September 1977.

7. G.R. Doddington, "Personal Identity Verification Using Voice," Electro 76 Professional Program, paper 22-4, May 11-14, 1976.

8. G.R. Doddington, R.E. Helms, and B.M. Hydrick, "Speaker Verification III," RADC-TR-76-262, August 1976, B014720L.

9. B.G. Secrest and R.E. Helms, "Remote Terminal Speaker Verification," RADC-TR-77-169, May 1977, A040827.

10. R.L. Davis, B.M. Hydrick, and G.R. Doddington, "Total Voice Speaker Verification," RADC-TR-78-260, January 1979, A065160.

11. Model 990 Computer DX10 Operating System Release 3 Reference Manual, Applications Programming Guide, Texas Instruments Incorporated, Part No. 946250-9703.

12. Model 990 Computer TX990 Operating System Documentation, Texas Instruments Incorporated, Manual No. 944776-9701.

13. TI Pascal Microprocessor Executive User's Guide, Texas Instruments Incorporated, Internal Publication.

14. A.R. Smith, B.P. Landell, and G. Vensko, "Solid State Audio/Speech Processor Analysis," RADC-TR-80-75, March 1980, A084743.

15. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. on ASSP, Vol. ASSP-23, February 1975, pp. 67-72.

16. J. LeRoux and C. Gueguen, "A Fixed Point Computation of Partial Correlation Coefficients," IEEE Trans. on ASSP, Vol. ASSP-25, June 1977, pp. 257-259.

17. A. Papoulis, Probability, Random Variables, and Stochastic Processes, New York: McGraw-Hill Book Co., 1965.

18. J.D. Markel, "Digital Inverse filtering - A New Tool for Formant Trajectory Estimation," IEEE Trans. on Audio and Electro-acoustics, Vol. AU-20, June 1972, pp. 129-137.

19. H. Wakita, "Residual Energy of Linear Prediciton Applied to Vowel and Speaker Recognition," IEEE Trans. on ASSP, Vol. ASSP-24, June 1976, pp. 270-271.

20. Segment Specification for Entry Control Subsystem, AFSC-ESD, Base and Installation Security System Program Office, Spec. No. BISS-ENC-14000, 15 August 1977.

21. T.H. Naylor, et al., Computer Simulation Techniques, New York: John Wiley and Sons, 1966.

22. N.A.J. Hastings and J.B. Peacock, Statistical Distributions, London: Butterworth, 1974.

23. M.J. Foodman, "Test Results - Advanced Development Models of BISS Identity Verification Equipment; Volume 2, Automatic Speaker Verification," Mitre Technical Report, MTR-3442, 1 September 1977.

Appendix I

ENTRY CONTROL POINT SIMULATION MODEL

Appendix I

ENTRY CONTROL POINT SIMULATION MODEL

There are two primary parts to the simulation model for an entry control point:

- o a user arrival rate model and
- o a booth service time model.

The simulation model developed in this study accounts for a variable number of booths at a single entry control point, arrivals of groups of users, and arrival rates that vary according to the time of day. Although allowing multiple people in the booth at one time is a desirable feature for improving throughput, only one occupant at a time was allowed as required in the BISS specification.[1]

1. User Arrival Rate Model

The queuing model used for user arrival rates is that groups of people arrive at times that are exponentially distributed. The exponential distribution of arrival times requires[2] that the probability that an arrival will occur in a small time interval is very small and that the occurrence of an arrival is statistically independent of the occurrence of other arrivals. This distribution is given by

$$f(t) = (1/b)\ \exp(-t/b) \quad ,$$

---

1. Segment Specification for Entry Control Subsystem, AFSC-ESD, Base and Installation Security System Program Office, Spec. No. BISS-ENC-14000, 15 August 1977.

2. T. H. Naylor, et al., Computer Simulation Techniques (John Wiley and Sons, New York, 1966).

where "b" is the mean arrival rate for groups of people. The need to model groups as exponentially distributed rather than individuals is that the arrival of an individual is not necessarily independent of other arrivals. This is obvious from observations of people leaving a common work area together or delaying their arrival by a few seconds in order to walk to the entry control point with other individuals. Since mean arrival rates in the BISS specification are in terms of people per minute, rather than groups per minute, "b" must be defined in terms of the probabilities of occurrence of various sized groups. Estimates of the probabilities of occurrence were obtained by measuring entries and exits of employees at the Semiconductor Building of Texas Instruments in Dallas. These probabilities are as follows:

n = number of people in group:	1	2	3	4	5	> 5
p(n) = p (number of people in group = n):	0.90	0.05	0.03	0.01	0.01	0

If "a" is the mean arrival rate for people per minute, then "b" is given by

$$b = a \Bigg/ \sum_{n=1}^{N_{max}} n * p(n) = a/1.18 \quad .$$

The mean arrival rates used in this model are those given in the BISS specification. These arrival rates (ARs) for entrants and extrants are reproduced in Figures 1 and 2, respectively, and are shown in tabular form in Table 1, where rates between table entries are found by linear interpolation. The BISS specification states, however, that only the activity over the period 0600 to 1800 hours should be used to demonstrate that the average throughput specification is met (thıee people per minute for voice authentication on both entrance and exit and four people per minute for voice authentication only on entrance) and that the maximum waiting time is not exceeded (three minutes at the 95th percentile). It should be noted, however, that these authors consider a maximum waiting time of three minutes at the 95th percentile to be sufficiently long to doom such an automated entry control system to failure

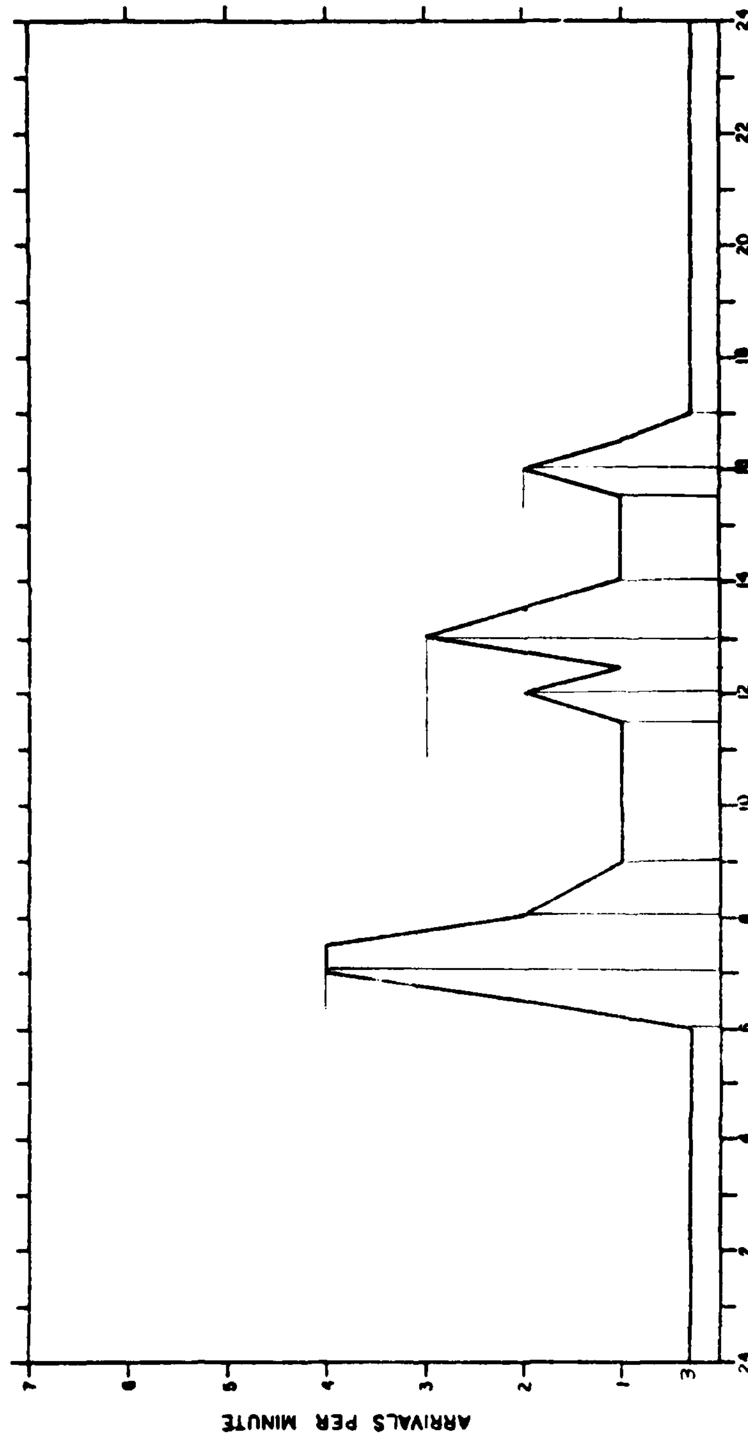


Figure 1 Arrival Rates for Entry

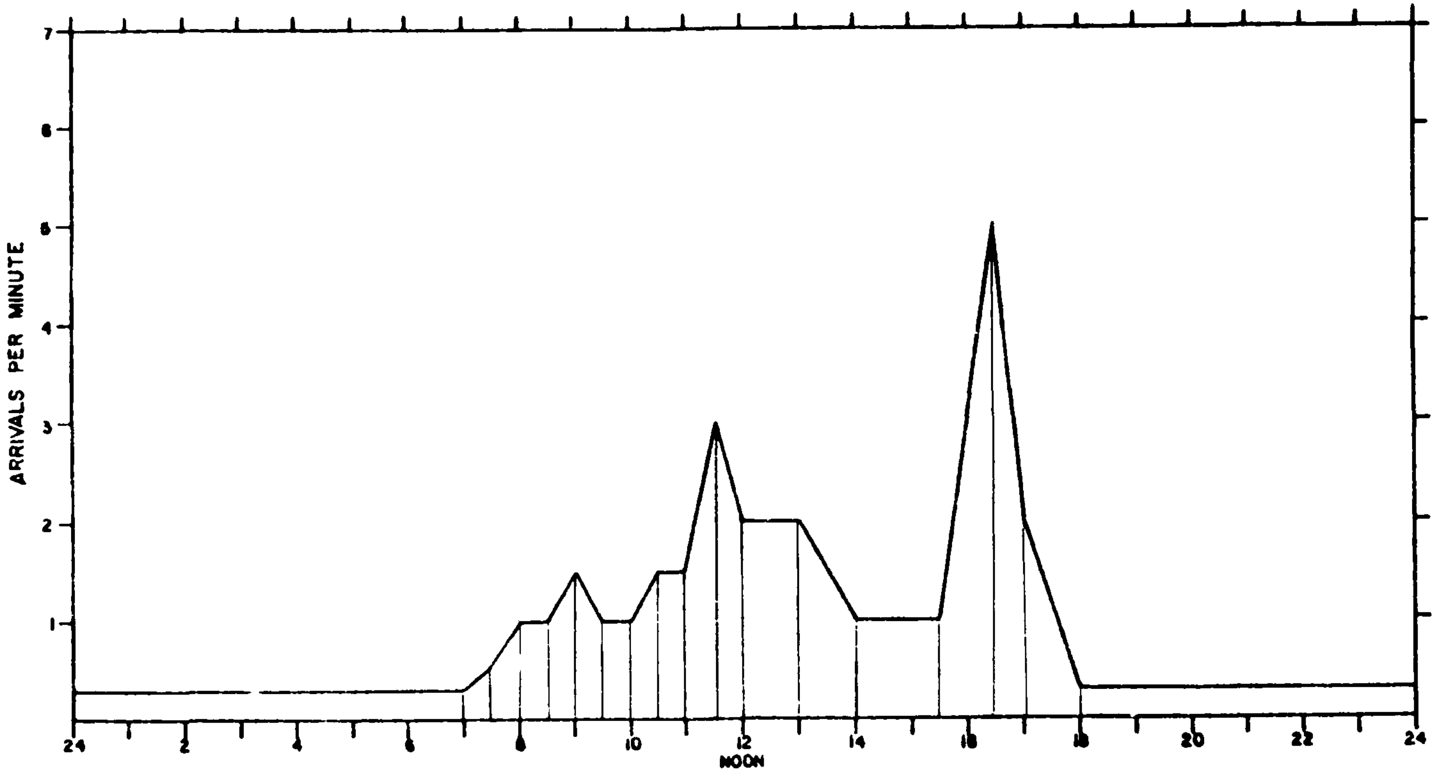


Figure 2 Arrival Rates for Exit

Table 1

Arrival Rates for Entrants and Extrants

Time Index	Time (Hours)	Arrival Rates (People/Min) Entrants	Extrants
0	0030 - 0530	0.3	0.3
1	0600	0.3	0.3
2	0630	2.15	0.3
3	0700	4.0	0.3
4	0730	4.0	0.5
5	0800	2.0	1.0
6	0830	1.5	1.0
7	0900	1.0	1.5
8	0930	1.0	1.0
9	1000	1.0	1.0
10	1030	1.0	1.5
11	1100	1.0	1.5
12	1130	1.0	3.0
13	1200	2.0	2.0
14	1230	1.0	2.0
15	1300	3.0	2.0
16	1330	2.0	1.5
17	1400	1.0	1.0
18	1430	1.0	1.0
19	1500	1.0	1.0
20	1530	1.0	1.0
21	1600	2.0	3.0
22	1630	1.0	5.0
23	1700	0.3	2.0
24	1730	0.3	1.15
25	1800 - 2400	0.3	0.3

due to lack of user acceptance. A maximum waiting time of 30 seconds at the 95th percentile would be more reasonable. Another important point is that the entry and exit rate profiles are themselves averages that vary according to day of the week, time of year, etc. Since using such averages in finding 95th or 98th percentile is imprudent at best, allowing such lax requirements as three minutes at the 95th percentile is certainly not appropriate.

The arrival rates (ARs) shown in Figures 1 and 2 and in Table 1 are for a specified population size. For populations of other sizes the entire entry and exit rate profiles must be multiplied by an amplitude factor (AF) as specified in Table 2.

Table 2

Entrant/Extrant Amplitude Factors for Various User Populations

User Population	Amplitude* Factor (AF)
500	0.37
1000	0.52
2000	0.80
2500	0.94
3000	1.09
4000	1.37
5000	1.66

* Derived from straight line fit for site 1 of bases A, B, and C; p. 172d of BISS specification. Equation for line is approximately AF = (number users/3500) + 0.23.

Hence, "b," the mean average arrival rate at a specified time "t" (in minutes) into a twelve hour day with 0600 hours corresponding to $t = 0$, is found by

$$b(t) = \left\{ AR(T) + [AR(T + 1) - AR(T)] * RATIO \right\} * AF/1.18 \quad ,$$

where RATIO = fractional part of t/30,

T = (integer part of t/30) + 1,

T + 1 = (T modulo 24) + 1,

and AR(T) is specified between the dashed lines in Table 1.

2. Booth Service Time Model

The booth service (or occupancy) time model is the second part to the model for an entry control point. This occupancy time is determined as the time required to open the module door, enter the module, allow the door to close, enter an ID number, verify your identity, and exit the booth allowing the door to close behind you. The model for this operation is given in the BISS specification, which describes the service time by a probability density function of the form:

$$P(t) = [(t - b)/a^2] \exp[-(t - b)/a] \quad , \tag{1}$$

where the service time parameters (a and b) are given in Table 3. These parameters were derived assuming (1) no verification on exit; and (2) that the ID entry device was located outside the booth, allowing the time for ID entry to be masked if the booth were occupied. Hence, the entries in Table 3 modeled the convolution of times for the events specified in Table 4, each of which was assumed to be exponentially distributed, except for the dead time.

Table 3

Booth Service Time Parameters

Type of Traffic	State of Pedestrian Module	Parameter Values (Seconds) a	b
Entry	Unoccupied	4.0	11.0
Entry	Occupied	3.5	9.5
Exit	Unoccupied	2.5	7.0
Exit	Occupied	2.0	5.5

Table 4

Components of Booth Service Time Model

Type of Traffic	State of Pedestrian Module	Components
Entry	Unoccupied	Two Doors, ID Entry, Verification, Dead Time
Entry	Occupied	Two Doors, Verification, Dead Time
Exit	Unoccupied	Two Doors, ID Entry, Dead Time
Exit	Occupied	Two Doors, Dead Time

Since the booth under consideration has the ID entry internal to the booth, only the first and third entries in Table 4 are of interest. There are other states, however, which need to be modeled that are not included in Tables 3 and 4. One set of these other states accounts for a "short" entry or exit occurring when a person or group is waiting to enter the booth by the same door a person is exiting by counting the door time only once. A flow chart showing the use of short transactions is given in Figure 3. (If multiple occupants were allowed, another set of states would be required to account for overlapping usage of the doors.) A revised list of the components of the booth service time model and the associated parameter values is given in Table 5. The new parameter values were derived using the same procedure as used for the original parameters that fit equation (1) to a more complicated expression derived using Laplace transforms. This booth service time model using Equation (1) and the parameters in Table 5 then reduce to generating random deviates that fit the equation. Equation (1) is an Erlang distribution for which random deviates may be generated[3] by

$$t = b - a \log \left[\prod_{i=1}^{2} r_i \right] ,$$

where r_i is a rectangular variate with range 0,1.

Using Table 5 as the basis for modeling the booth also allows the verification to be modeled separately and added to the booth time derived using the second and fourth entries in Table 5. The necessity for modeling the booth separately arises since user rejections have been neglected in the BISS specification model. To account for this, a model for verification has been established in which the time for each phrase to be said is modeled with

3. N.A.J. Hastings and J. B. Peacock, Statistical Distributions (Butterworth, London, 1974).

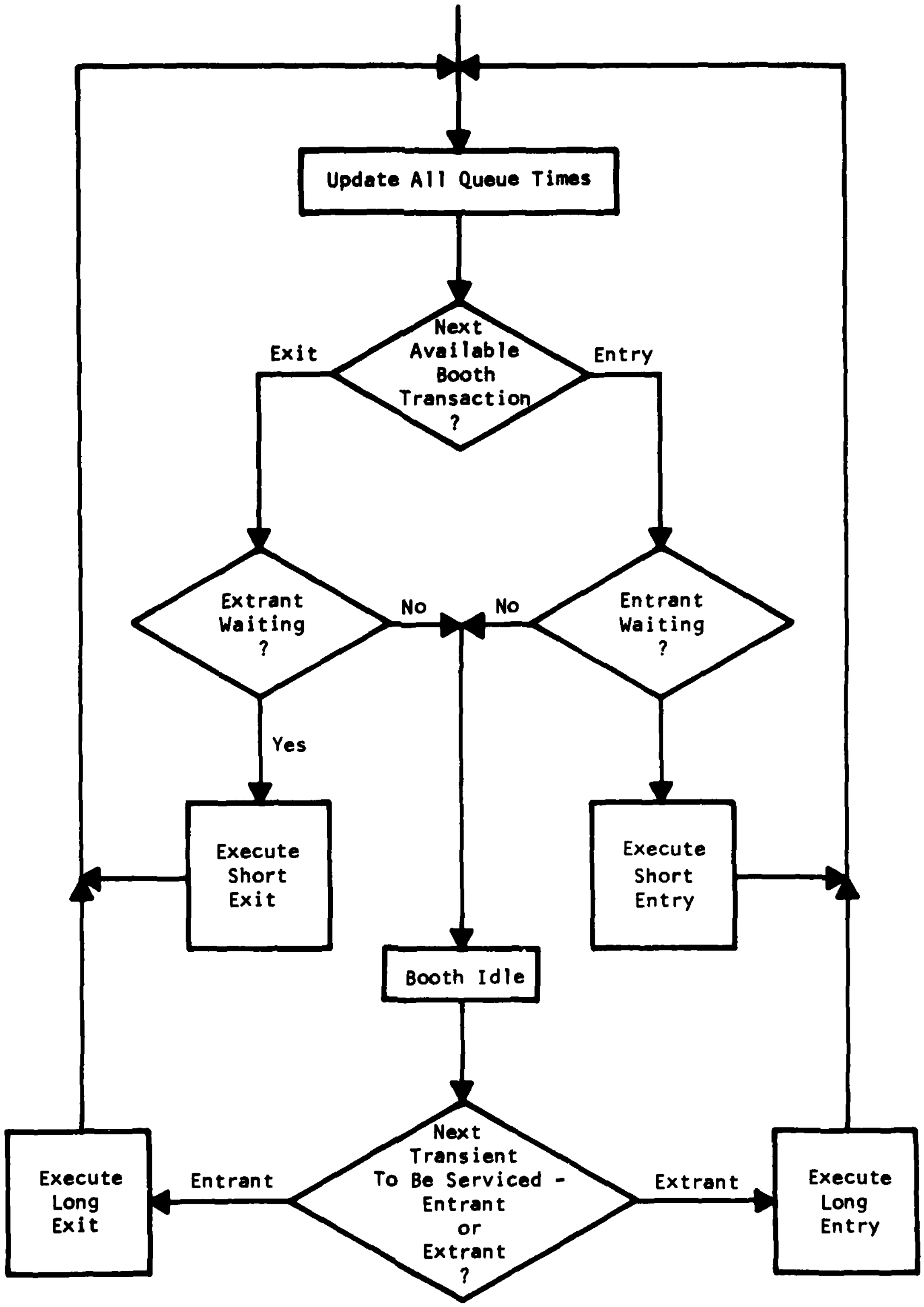


Figure 3 Overall Flow Diagram for Entry Control Point Simulation Model

Table 5

Revised Booth Service Time Parameters

Type of Transaction	Verification Required	Components	Parameter Values (Seconds) a	b
Long	Yes	Two Doors, ID Entry, Verification, Dead Time	4.0	11.0
Long	No	Two Doors, ID Entry, Dead Time	2.5	7.0
Short	Yes	One Door, ID Entry, Verification, Dead Time	3.15	8.5
Short	No	One Door, ID Entry, Dead Time	1.35	4.5

Table 6

Cumulative Verification Probabilities

Number of Phrases: N	P (Verified in N Phrases or Less)
1	0.613
2	0.864
3	0.946
4	0.970
5	0.973
6	0.975
7	0.979
8	0.987
9	0.990
10	0.990
11	0.990
12 (N_{Max})	0.990

a fourth order χ^2 distribution, and a "not verified" decision was accompanied by its own fourth order χ^2 distribution. The number of phrases, N, required for verification was determined such that p (verified in N - 1 phrases or less) $< r \leq$ p (verified in N phrases or less), where r is a rectangular variate with range 0,1. Obviously, $r >$ p (verified in NMAX phrases or less) is the "not verified" case. The probabilities used in this model (shown in Table 6) were taken from Table LIV of Volume 2 of the BISS test results,[4] assuming 1% not verified.

The χ^2 variable for a phrase prompt and response must be translated and scaled such that it is of the form

$$\frac{\left[\frac{t - t_{min}(ph)}{a}\right] \exp \left\{-\frac{[t - t_{min}(ph)]}{a}\right\}}{2} \qquad (2)$$

For χ^2 variables, the mean is ν (= 4 in this case), and the standard deviation is $\sqrt{2\nu}$. Using an average prompting time of 1.9 s, average response time of 1.93 s, and standard deviation of response time of 0.4 s, (p. 129 of reference 4), we have

$$4 = [t_{ave} - t_{min}(ph)] \Big/ a = [3.83 - t_{min}(ph)]/a$$

and $\sqrt{8}$ = stand. dev./a = $0.4(4)/[3.83 - t_{min}(ph)]$

yielding $t_{min}(ph) = 3.26$

$$a = 0.1425 = \frac{t_{ave}(ph) - t_{min}(ph)}{4}$$

4. Martin J. Foodman, "Test Results - Advanced Development Models of BISS Identity Verification Equipment; Volume 2, Automatic Speaker Verification," Mitre Technical Report MTR-3442, 1 September 1977.

The χ^2 variable for the "not verified" model must also be translated and scaled as in Equation (2). From p. 129 of Reference 4, the average verification time is given as 6.2 s. The average time allowed in the BISS model for verification, however, is just the difference in means (b + 2a) between the first two entries in Table 5, which is 7 s. With a 99% verification rate, the average for "not verifieds" can be found by

$$0.99\ (6.2) + 0.01 \left[t_{ave}\ (nv)\right] = 7.0$$

$$t_{ave}\ (nv) = 86.2\ s.$$

Letting t_{min} for "not verified" equal twelve times t_{min} for each phrase yields t_{min} (nv) $\approx$ 40 s and a = 11.55.

Hence, the models for verification times are as follows:

$$t_v = \sum_{i=1}^{N} (3.26 + 0.1425\ \chi_i)$$

$$t_{nv} = 40.0 + 11.55\ \chi \quad ,$$

where the "χ's" are chi-squared random deviates generated by

$$\chi = 2\ \log \left[\prod_{i=1}^{(\nu-1)/2} r_i\right] + n^2 \quad ,$$

where r_i is a rectangular variate with range 0,1 and n is a normal variate with mean 0 and standard deviation 1.

In conclusion it should be noted that the "dead" time does not include time lost due to insufficient processing capability. The assumption has been

made in this section that sufficient computing power is available to provide the user immediate response.

3. <u>Results</u>

Simulations were run using the entry and exit rate profiles shown in Figures 1 and 2 for the seven amplitude factors (AF) given in Table 2 and for AF = 1.0. Simulations were run for each of the following cases, where each run simulated five hundred, 12-hour days:

1. Number of booths varying from a minimum number (determined below) through six.
2. Verification required (a) on entry alone and (b) on both entry and exit.
3. Verification modeled (a) as in the BISS specification and (b) using the chi-squared model for each verification phrase.

The minimum number of booths can be determined from

NB > total processing time/elapsed time

NB > AF [(ave entrant rate)(ave entrant processing time)
+ (ave extrant rate) (ave extrant processing time)].

Assuming all transactions are "long" ones, the average (b + 2a) booth times for the first two entries in Table 5 are 19 and 12 respectively.

Figures 1 and 2 show average entrant and extrant rate profiles. Although the average rates specified in the above inequality could be the average of these profiles over the 0600 - 1800 time period (1.48 people/min. for both entrants and extrants), since these profiles are themselves averages it is more appropriate to use the maxima of the average profiles (4 entrants/min and 5 extrants/min). Maximum AFs as a function of various NB for both assumptions are given in Table 7.

Table 7

Maximum Amplitude Factors (AFs) as a Function Of Number of Booths (NB)

	Maximum AF	
Condition	Using Maximum of Average Profiles	Using Average of Average Profiles
Verification on Entry and Exit	0.35 NB	1.07 NB
Verification on Entry Only	0.44 NB	1.31 NB

Selected comparisons of these simulations are made in Figures 4 - 8. Note that all of these figures except Figure 5 are on log-by-probability scales, where all plots show the probability that a specified time is less than a certain value. Figure 5 is the same as Figure 4 except the curves are plotted on linear-by-probability scales instead.

Figures 4 and 5 show the waiting time for an available booth for various amplitude factors, for the case of three booths, using the TI modeled verification, with verification on both entry and exit.

Figure 6 shows not only the waiting time for an available booth, but the total delay through the booth (waiting and processing) as well. Both times are plotted for a varying number of booths, for an AF of one, using the TI modeled verification, with verification on both entry and exit. The inflections in the total delay curves around the 99th percentile are due to the 1% "not verifieds" which have much longer average processing times than for successful verifications.

The fact that "not verifieds" are not explicitly accounted for in the BISS specification (the MITRE model) becomes apparent in Figure 7. Figure 7 compares the MITRE model to the TI model for the case of five booths, an AF of one, with verification on both entry and exit.

Figure 8 shows the advantage of verification only on entry if the scenario permits. This figure shows both the total delay and just the waiting time for the case of five booths, an AF of one, and using the TI modeled verification both for entry alone and for both entry and exit.

More details of the simulations are given in Tables 8 - 11 which give the averages in seconds of both the waiting time alone and the waiting plus processing time for both entry and exits, and in Tables 12 - 15 which give the 95th and 98th percentiles in seconds for both the waiting time and the total delay. The increased throughput shown in Tables 8 - 11 for the longer waiting

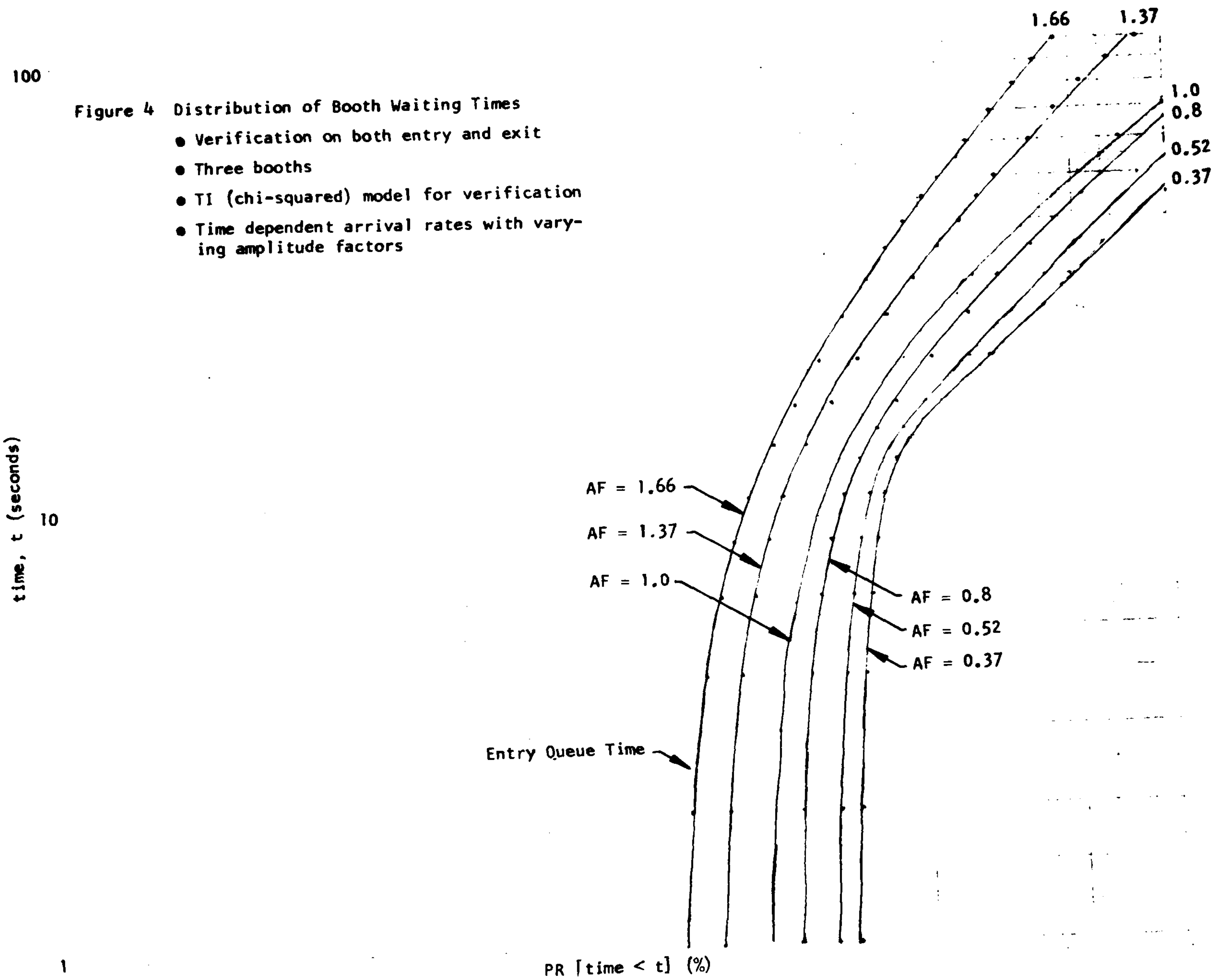
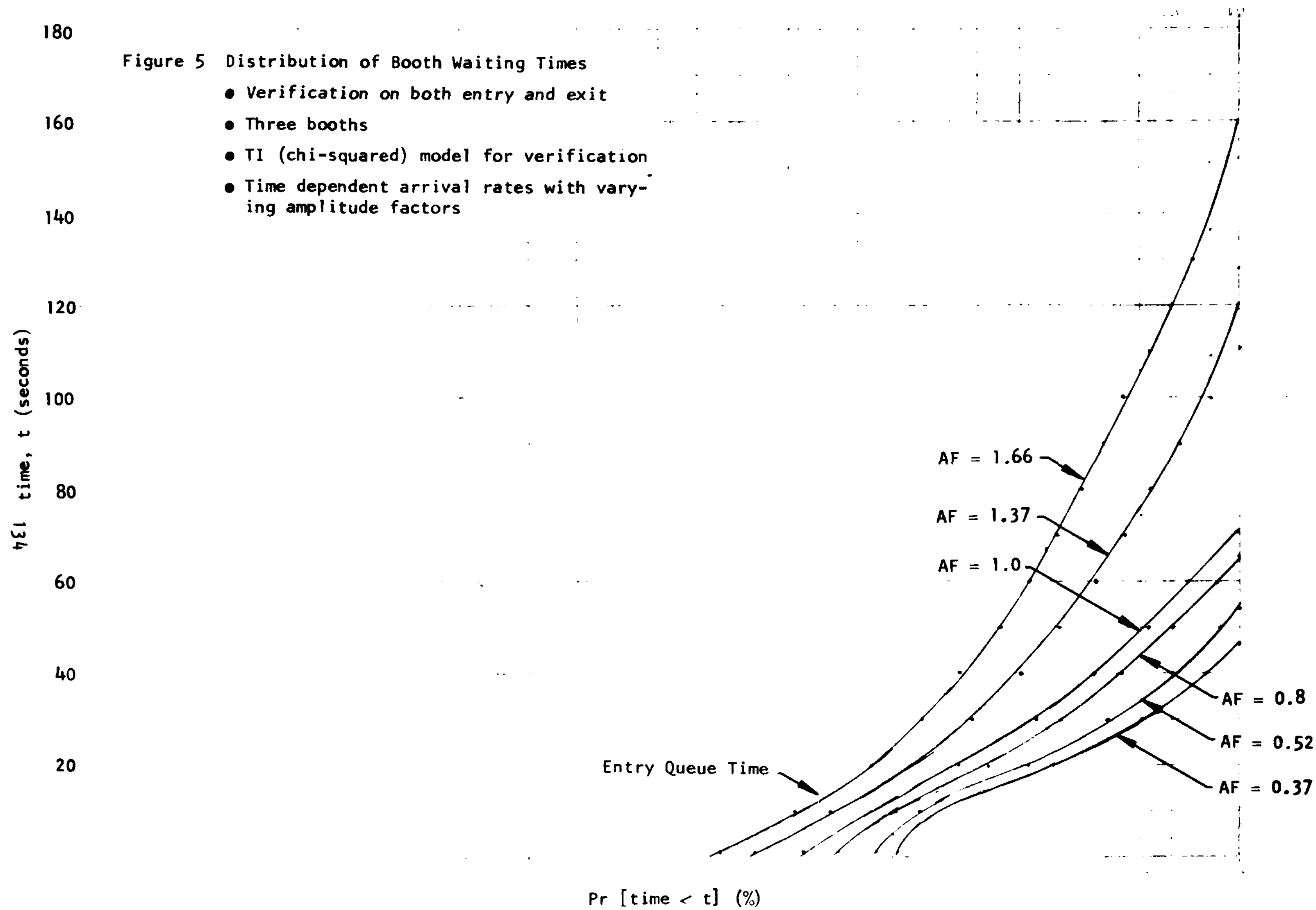


Figure 4 Distribution of Booth Waiting Times

- Verification on both entry and exit
- Three booths
- TI (chi-squared) model for verification
- Time dependent arrival rates with varying amplitude factors

Figure 5 Distribution of Booth Waiting Times

- Verification on both entry and exit
- Three booths
- TI (chi-squared) model for verification
- Time dependent arrival rates with varying amplitude factors



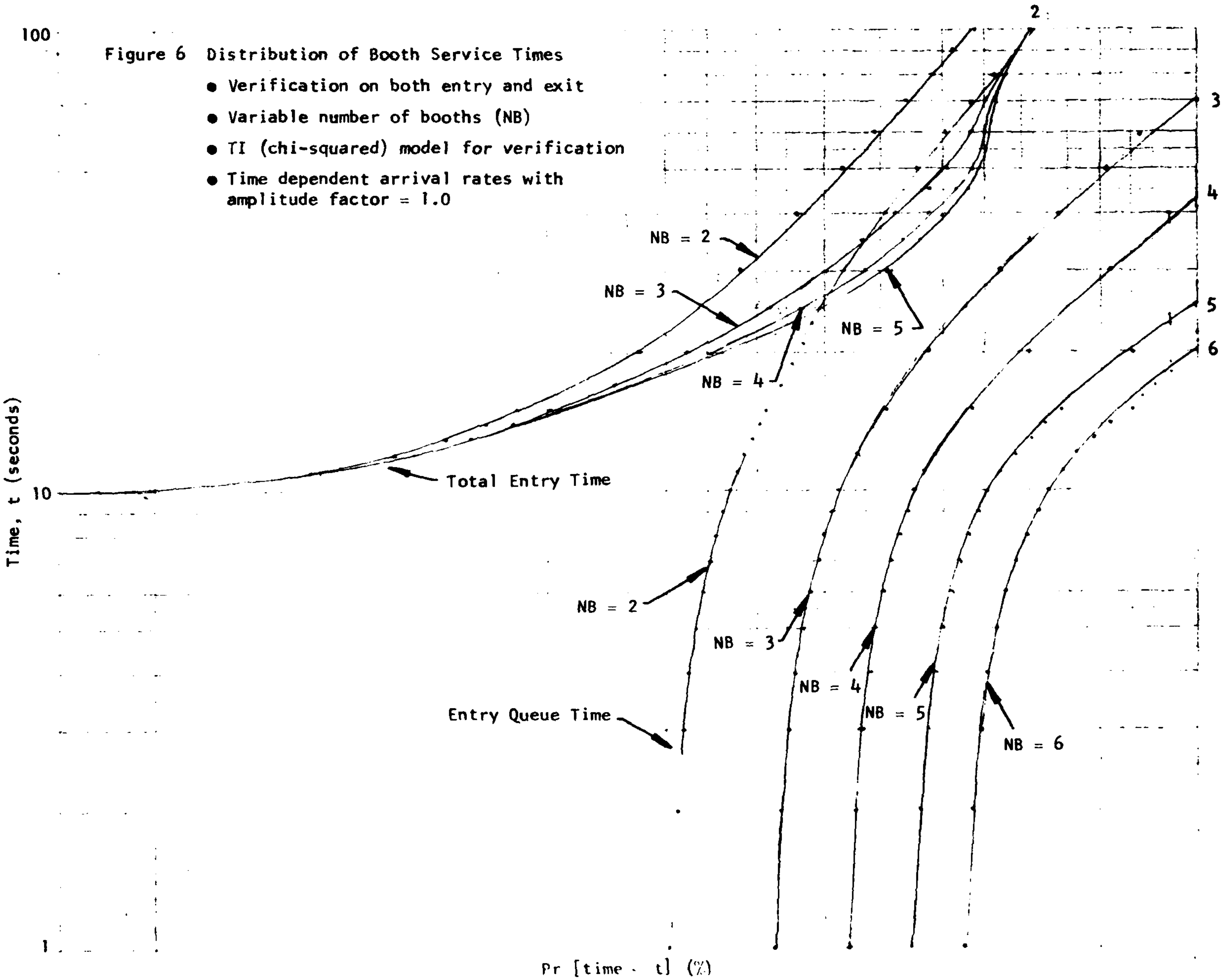


Figure 6 Distribution of Booth Service Times

- Verification on both entry and exit
- Variable number of booths (NB)
- TI (chi-squared) model for verification
- Time dependent arrival rates with amplitude factor = 1.0

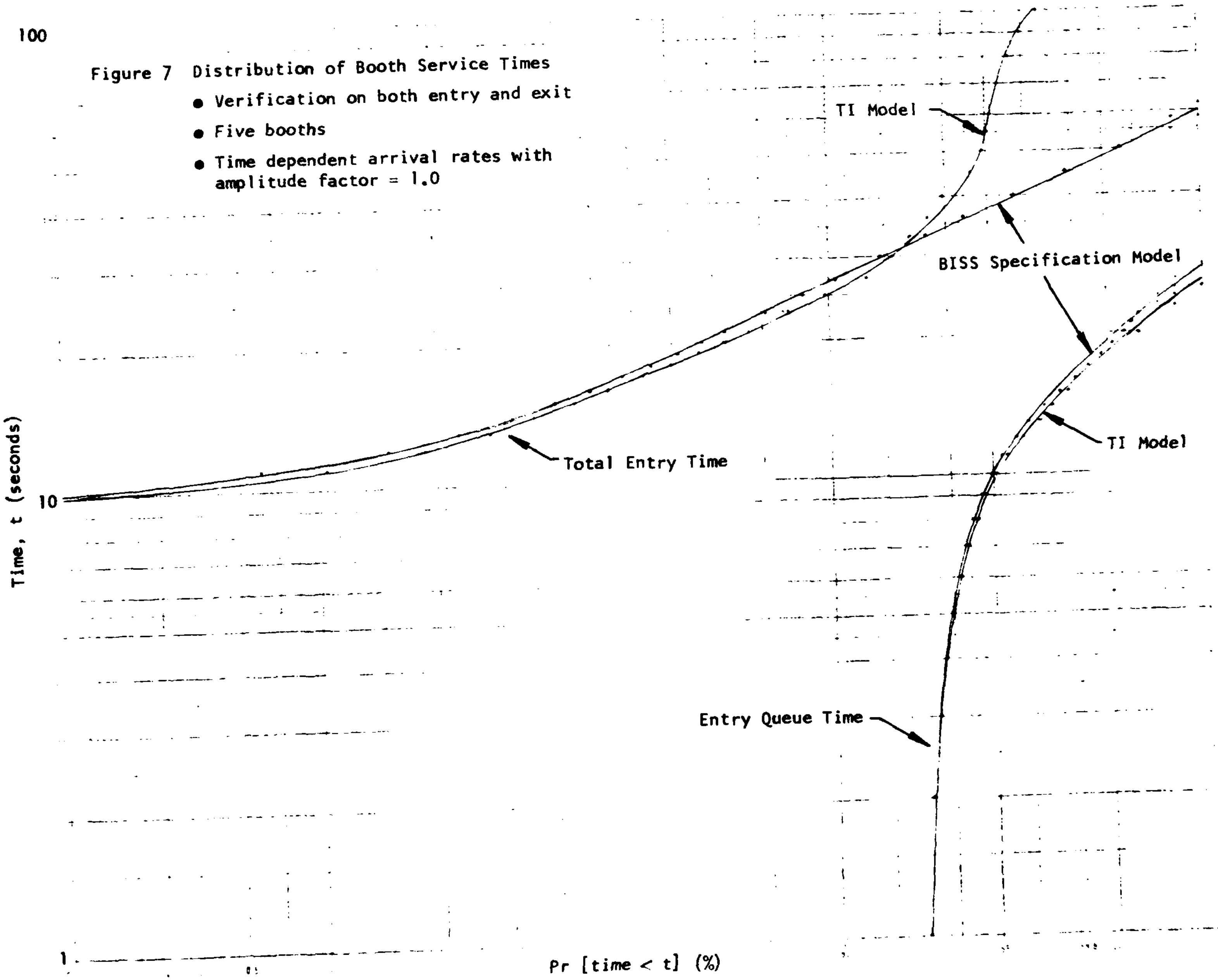


Figure 7 Distribution of Booth Service Times

- Verification on both entry and exit
- Five booths
- Time dependent arrival rates with amplitude factor = 1.0

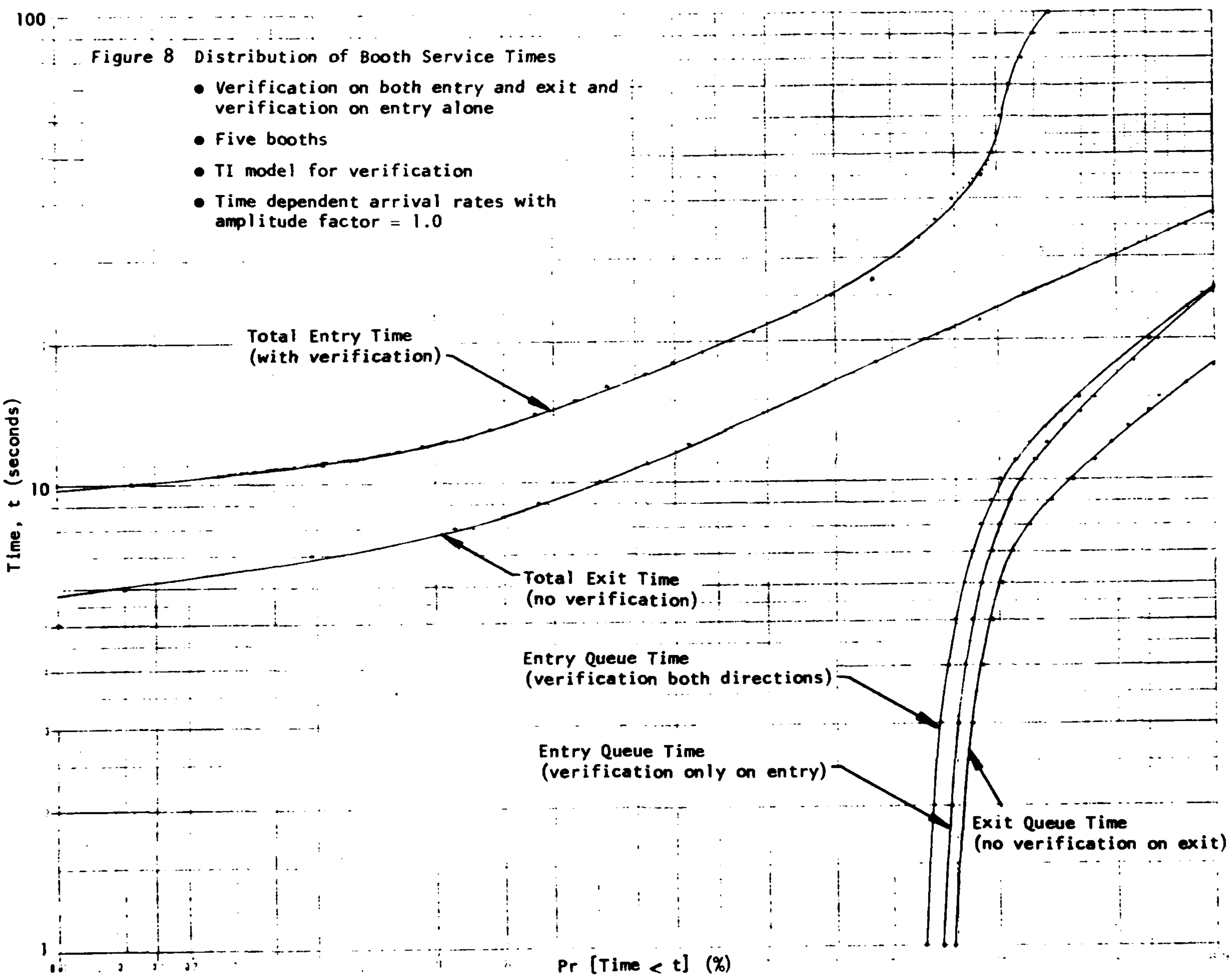


Figure 8 Distribution of Booth Service Times

- Verification on both entry and exit and verification on entry alone
- Five booths
- TI model for verification
- Time dependent arrival rates with amplitude factor = 1.0

Table 8

Average Waiting Times and Total Delays (in Seconds) Using TI-Modeled Verification on Both Entry and Exit

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
0.37	1	13.8	25.9	32.1	27.5	14.7	28.7	32.9	30.2	3.29
↓	2	2.5	7.3	21.3	12.6	2.6	7.6	21.4	12.8	3.19
	3	0.8	3.5	19.7	10.9	0.8	3.5	19.7	10.7	3.17
	4	0.3	1.9	19.2	10.4	0.3	1.8	19.2	10.3	3.16
	5	0.1	0.9	19.0	10.3	0.1	0.8	19.0	10.3	3.16
0.37	6	0.0	0.4	19.0	10.2	0.0	0.4	19.0	10.2	3.16
0.52	1	21.5	36.9	39.4	37.9	23.1	41.6	41.0	42.5	3.35
↓	2	3.3	8.7	22.0	13.2	3.4	9.0	22.1	13.5	3.21
	3	1.0	4.0	19.9	10.9	1.0	4.0	19.9	10.8	3.18
	4	0.3	2.1	19.3	10.4	0.3	2.1	19.3	10.3	3.17
	5	0.1	1.0	19.1	10.5	0.1	1.0	19.1	10.3	3.16
0.52	6	0.0	0.5	19.0	10.3	0.0	0.5	19.0	10.4	3.16
0.8	1	81.7	141.5	98.9	142.0	103.0	187.1	120.2	187.5	3.48
↓	2	5.6	12.3	24.0	15.7	5.8	13.1	24.2	16.3	3.26
	3	1.6	5.2	20.4	11.3	1.6	5.2	20.3	11.4	3.20
	4	0.5	2.6	19.4	10.5	0.6	2.7	19.4	10.5	3.18
	5	0.2	1.4	19.1	10.3	0.2	1.4	19.2	10.3	3.16
0.8	6	0.1	0.7	19.0	10.3	0.1	0.8	19.0	10.2	3.16
0.94	2	7.4	15.1	25.6	18.0	8.0	17.5	26.3	20.0	3.29
↓	3	1.9	5.7	20.6	11.5	2.0	5.9	20.7	11.7	3.21
	4	0.7	3.0	19.5	10.5	0.7	3.1	19.6	10.7	3.18

Table 8

Average Waiting Times and Total Delays (in Seconds) Using TI-Modeled Verification on Both Entry and Exit

(Continued)

AF	NB	Entrance Wait Avg (sec)	Entrance Wait STDEV	Entrance Wait + Proc Avg (sec)	Entrance Wait + Proc STDEV	Exit Wait Avg (sec)	Exit Wait STDEV	Exit Wait + Proc Avg (sec)	Exit Wait + Proc STDEV	Avg Throughput (people/min)
	5	0.2	1.6	19.2	10.4	0.3	1.7	19.2	10.4	3.17
0.94	6	0.1	0.9	19.1	10.4	0.1	0.9	19.0	10.2	3.16
1.00	2	8.2	16.5	26.3	19.1	9.2	19.7	27.4	21.9	3.30
	3	2.1	6.0	20.8	11.7	2.2	6.4	20.9	12.0	3.21
	4	0.8	3.2	19.6	10.7	0.8	3.3	19.6	10.7	3.18
	5	0.3	1.7	19.2	10.3	0.3	1.7	19.2	10.5	3.17
1.00	6	0.1	1.0	19.1	10.3	0.1	1.0	19.1	10.2	3.16
1.09	2	9.7	18.7	27.8	21.0	11.6	24.7	29.7	26.5	3.32
	3	2.4	6.5	21.0	11.8	2.5	6.9	21.1	12.1	3.23
	4	0.9	3.3	19.7	10.8	0.9	3.5	19.7	10.8	3.18
	5	0.3	1.8	19.2	10.4	0.3	1.9	19.3	10.4	3.17
1.09	6	0.1	1.0	19.1	10.3	0.1	1.0	19.1	10.2	3.16
1.37	2	20.2	38.5	37.9	39.7	29.5	64.1	47.2	64.9	3.39
	3	3.7	8.8	22.2	13.3	4.1	10.3	22.5	14.2	3.26
	4	1.3	4.1	20.0	10.9	1.4	4.5	20.1	11.0	3.21
	5	0.5	2.3	19.3	10.5	0.5	2.4	19.4	10.5	3.18
1.37	6	0.2	1.3	19.1	10.3	0.2	1.4	19.1	10.2	3.17
1.66	2	63.1	117.6	80.4	118.1	89.2	177.5	106.6	177.8	3.46
	3	5.7	12.2	23.9	15.6	7.5	18.6	25.7	20.9	3.30
	4	1.9	5.3	20.5	11.3	2.0	5.8	20.6	11.7	3.22
	5	0.7	2.8	19.5	10.6	0.8	3.1	19.6	10.7	3.19
1.66	6	0.3	1.7	19.2	10.3	0.3	1.8	19.2	10.4	3.17

Table 9

Average Waiting Times and Total Delays (in Seconds) Using BISS Specification Model for Verification on Both Entry and Exit

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
0.37	1	13.2	23.2	31.5	23.6	13.7	24.5	32.0	24.9	3.28
↓	2	2.6	7.4	21.4	9.2	2.5	7.3	21.4	9.1	3.19
	3	0.8	3.5	19.7	6.6	0.8	3.6	19.8	6.6	3.17
	4	0.3	1.9	19.2	5.9	0.3	1.9	19.2	5.9	3.16
	5	0.1	0.9	19.1	5.7	0.1	0.9	19.1	5.7	3.16
0.37	6	0.0	0.4	19.0	5.7	0.0	0.4	19.0	5.7	3.16
0.52	1	20.3	33.2	38.4	33.4	22.1	38.8	40.1	38.9	3.33
↓	2	3.3	8.5	22.0	10.0	3.4	8.7	22.1	10.1	3.21
	3	1.0	4.0	19.9	6.8	1.0	4.0	19.9	6.8	3.18
	4	0.4	2.2	19.3	6.0	0.4	2.2	19.3	6.0	3.16
	5	0.1	1.0	19.1	5.7	0.1	1.0	19.1	5.7	3.16
0.52	6	0.0	0.5	19.0	5.7	0.0	0.5	19.0	5.7	3.16
0.80	1	79.6	135.0	97.0	135.2	106.6	192.0	124.0	192.2	3.44
↓	2	5.4	11.3	23.9	12.3	5.7	12.2	24.1	13.1	3.25
	3	1.6	5.1	20.4	7.4	1.7	5.3	20.5	7.5	3.19
	4	0.6	2.7	19.5	6.2	0.6	2.8	19.5	6.2	3.17
	5	0.2	1.5	19.1	5.8	0.2	1.5	19.2	5.8	3.17
0.80	6	0.1	0.8	19.1	5.7	0.1	0.8	19.0	5.7	3.16
0.94	2	7.1	13.7	25.4	14.4	7.6	15.7	26.0	16.3	3.27
↓	3	2.0	5.7	20.7	7.8	2.0	5.9	20.8	7.9	3.20
	4	0.7	3.0	19.6	6.3	0.7	3.1	19.6	6.3	3.18

Table 9

BISS Specification Model for Verification on Both Entry and Exit
(Continued)

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
	5	0.2	1.6	19.2	5.8	0.3	1.7	19.2	5.8	3.17
0.94	6	0.1	0.9	19.1	5.7	0.1	1.0	19.1	5.7	3.16
1.0	2	7.8	14.8	26.0	15.5	8.9	18.6	27.2	19.1	3.29
	3	2.1	6.0	20.9	8.0	2.2	6.3	21.0	8.2	3.21
	4	0.8	3.2	19.6	6.4	0.8	3.3	19.7	6.4	3.18
	5	0.3	1.8	19.2	5.9	0.3	1.8	19.2	5.9	3.17
1.0	6	0.1	0.9	19.1	5.7	0.1	1.1	19.1	5.7	3.16
1.00	2	9.3	17.2	27.5	17.7	11.6	24.6	29.7	25.0	3.30
	3	2.4	6.4	21.1	8.2	2.5	6.8	21.2	8.5	3.22
	4	0.9	3.4	19.7	6.5	0.9	3.5	19.7	6.5	3.19
	5	0.3	1.9	19.3	5.9	0.3	2.0	19.3	5.9	3.17
1.09	6	0.1	1.1	19.1	5.7	0.1	1.1	19.1	5.7	3.16
1.37	2	20.1	37.3	37.9	37.7	30.1	64.8	47.9	65.0	3.36
	3	3.7	8.3	22.2	9.7	4.1	9.6	22.6	10.7	3.25
	4	1.3	4.1	20.0	6.8	1.4	4.5	20.1	7.0	3.20
	5	0.5	2.3	19.4	6.0	0.5	2.4	19.4	6.1	3.18
1.37	6	0.2	1.4	19.1	5.8	0.2	1.5	19.2	5.8	3.17
1.66	2	66.1	120.9	83.6	121.1	93.6	182.9	111.1	183.0	3.43
	3	5.7	11.5	23.9	12.5	7.4	17.3	25.6	17.9	3.28
	4	1.9	5.2	20.5	7.4	2.1	6.0	20.7	7.9	3.22
	5	0.7	2.9	19.5	6.2	0.8	3.1	19.6	6.3	3.19
	6	0.3	1.7	19.2	5.8	0.3	1.8	19.2	5.9	3.17

Table 10

Average Waiting Times and Total Delays (in Seconds) Using

TI-Modeled Verification on Entry Only

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
0.37	1	11.5	22.8	30.0	24.9	7.3	14.6	18.7	14.5	4.03
	2	2.2	6.8	21.1	12.3	1.5	4.6	13.3	5.6	3.91
	3	0.7	3.2	19.6	10.7	0.5	2.2	12.4	4.1	3.88
	4	0.2	1.7	19.2	10.2	0.2	1.2	12.1	3.7	3.88
	5	0.1	0.7	19.1	10.5	0.0	0.5	12.0	3.5	3.87
0.37	6	0.0	0.3	19.0	10.1	0.0	0.2	12.0	3.5	3.88
0.52	1	17.0	31.2	35.2	32.7	9.9	18.1	21.0	17.9	4.10
	2	2.8	7.8	21.6	12.7	1.9	5.2	13.6	6.0	3.93
	3	0.8	3.6	19.8	10.7	0.6	2.5	12.5	4.2	3.89
	4	0.3	1.9	19.3	10.4	0.2	1.3	12.2	3.7	3.88
	5	0.1	0.8	19.1	10.3	0.1	0.6	12.0	3.6	3.87
0.52	6	0.0	0.4	19.0	10.2	0.0	0.3	12.0	3.5	3.87
0.8	1	54.3	104.0	72.0	104.7	19.9	34.5	30.4	34.3	4.25
	2	4.4	10.5	23.0	14.6	2.8	6.6	14.3	7.0	3.98
	3	1.2	4.4	20.1	11.0	0.9	3.1	12.7	4.5	3.91
	4	0.4	2.3	19.4	10.5	0.3	1.6	12.2	3.8	3.89
	5	0.1	1.2	19.1	10.3	0.1	0.9	12.1	3.6	3.88
0.8	6	0.0	0.6	19.0	10.1	0.0	0.5	12.0	3.6	3.88
0.94	1	126.4	231.9	143.8	232.2	33.7	60.7	44.0	60.6	4.34
	2	5.4	12.1	23.9	15.5	3.3	7.5	14.7	7.8	4.02
	3	1.5	4.9	20.3	11.3	1.0	3.4	12.8	4.6	3.92

Table 10

Average Waiting Times and Total Delays (in Seconds) Using

TI-Modeled Verification on Entry Only

(Continued)

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
↓	4	0.5	2.6	19.5	10.6	0.4	1.8	12.3	3.9	3.89
	5	0.2	1.3	19.1	10.3	0.1	1.0	12.1	3.6	3.88
0.94	6	0.1	0.7	19.0	10.2	0.0	0.5	12.0	3.6	3.88
1.0	1	171.9	297.8	189.1	298.0	41.7	77.7	51.9	77.7	4.37
↓	2	6.1	13.5	24.6	16.7	3.6	7.9	15.0	8.2	4.02
	3	1.6	5.2	20.4	11.4	1.1	3.5	12.8	4.7	3.93
	4	0.6	2.7	19.5	10.6	0.4	1.8	12.3	3.9	3.90
	5	0.2	1.4	19.2	10.4	0.1	1.1	12.1	3.6	3.88
1.0	6	0.1	0.7	19.1	10.3	0.0	0.6	12.0	3.6	3.87
1.09	1	252.3	409.8	269.5	410.0	61.6	114.9	71.6	115.1	4.43
↓	2	7.3	15.7	25.6	18.6	4.1	8.4	15.3	8.6	4.05
	3	1.8	5.5	20.6	11.5	1.2	3.7	12.9	4.9	3.94
	4	0.6	2.9	19.5	10.6	0.4	2.0	12.3	3.9	3.90
	5	0.2	1.5	19.2	10.3	0.2	1.1	12.1	3.7	3.89
1.09	6	0.1	0.8	19.1	10.4	0.1	0.6	12.0	3.6	3.87
1.37	2	14.3	31.2	32.4	32.7	6.0	11.5	17.0	11.5	4.12
↓	3	2.8	7.4	21.4	12.6	1.7	4.5	13.2	5.4	3.97
	4	0.9	3.5	19.7	10.7	0.6	2.4	12.4	4.1	3.92
	5	0.3	1.9	19.2	10.3	0.2	1.4	12.1	3.7	3.89
1.37	6	0.1	1.1	19.1	10.2	0.1	0.8	12.1	3.6	3.88

Table 10

Average Waiting Times and Total Delays (in Seconds) Using
TI-Modeled Verification on Entry Only
(Continued)

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
1.66	2	43.2	94.4	61.0	95.1	10.2	20.1	20.9	20.0	4.21
	3	4.1	9.8	22.5	14.1	2.2	5.4	13.7	6.0	4.01
	4	1.3	4.4	20.1	11.0	0.8	2.8	12.6	4.3	3.94
	5	0.5	2.4	19.4	10.5	0.3	1.6	12.2	3.8	3.90
1.66	6	0.2	1.3	19.1	10.3	0.1	1.0	12.1	3.6	3.89

Table 11

Average Waiting Times and Total Delays (in Seconds) Using BISS Specification Model for Verification on Entry Only

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
0.37	1	11.3	21.0	29.8	21.6	7.0	13.1	18.4	13.2	4.02
	2	2.2	6.8	21.1	8.7	1.5	4.5	13.3	5.5	3.91
	3	0.7	3.3	19.7	6.5	0.5	2.2	12.4	4.1	3.88
	4	0.2	1.8	19.2	5.9	0.1	1.2	12.1	3.7	3.88
	5	0.1	0.7	19.1	5.7	0.0	0.6	12.0	3.6	3.87
0.37	6	0.0	0.3	19.0	5.7	0.0	0.3	12.0	3.5	3.87
0.52	1	16.1	28.2	34.4	28.6	9.6	16.3	20.7	16.2	4.09
	2	2.8	7.7	21.6	9.4	1.9	5.1	13.6	6.0	3.93
	3	0.9	3.7	19.8	6.7	0.6	2.5	12.5	4.2	3.89
	4	0.3	2.0	19.3	6.0	0.2	1.4	12.2	3.7	3.88
	5	0.1	0.9	19.1	5.7	0.1	0.7	12.0	3.6	3.87
0.52	6	0.0	0.4	19.0	5.7	0.0	0.3	12.0	3.6	3.87
0.8	1	51.3	98.4	69.2	98.6	18.8	31.0	29.4	30.9	4.22
	2	4.2	9.7	22.8	11.0	2.7	6.4	14.3	6.9	3.98
	3	1.2	4.4	20.1	7.1	0.9	3.1	12.7	4.5	3.91
	4	0.4	2.4	19.4	6.1	0.3	1.6	12.2	3.8	3.89
	5	0.1	1.2	19.1	5.8	0.1	0.9	12.1	3.6	3.88
0.8	6	0.0	0.6	19.0	5.7	0.0	0.5	12.0	3.5	3.87
0.94	1	122.9	221.3	140.5	221.4	33.3	60.6	43.5	60.6	4.31
	2	5.3	11.4	23.8	12.5	3.3	7.2	14.7	7.5	4.01
	3	1.5	5.0	20.4	7.4	1.0	3.3	12.8	4.6	3.92
	4	0.5	2.6	19.5	6.2	0.4	1.8	12.3	3.9	3.89

Table 11

Average Waiting Times and Total Delays (in Seconds) Using BISS Specification Model for Verification on Entry Only (Continued)

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
↓	5	0.2	1.4	19.1	5.8	0.1	1.0	12.1	3.6	3.88
0.94	6	0.1	0.8	19.1	5.7	0.0	0.6	12.0	3.6	3.87
1.0	1	171.3	296.3	188.8	296.4	42.2	78.6	52.4	78.6	4.34
	2	5.9	12.5	24.4	13.5	3.6	7.5	14.9	7.8	4.02
	3	1.7	5.3	20.5	7.6	1.1	3.5	12.8	4.7	3.93
	4	0.6	2.7	19.5	6.2	0.4	1.9	12.3	3.9	3.89
↓	5	0.2	1.4	19.2	5.8	0.1	1.1	12.1	3.6	3.88
1.0	6	0.1	0.8	19.1	5.7	0.0	0.6	12.0	3.6	3.88
1.09	1	263.1	418.0	280.4	418.1	65.2	119.0	75.2	119.2	4.40
	2	7.0	14.5	25.4	15.3	4.0	8.0	15.3	8.2	4.04
	3	1.9	5.5	20.6	7.7	1.2	3.7	12.9	4.8	3.94
	4	0.7	2.9	19.6	6.3	0.4	2.0	12.3	3.9	3.90
↓	5	0.2	1.5	19.2	5.8	0.2	1.1	12.1	3.7	3.88
1.09	6	0.1	0.8	19.1	5.7	0.1	0.7	12.0	3.6	3.88
1.37	2	13.2	27.4	31.4	27.9	5.9	11.0	16.9	11.0	4.11
	3	2.7	6.9	21.3	8.7	1.6	4.4	13.2	5.2	3.97
	4	0.9	3.5	19.8	6.5	0.6	2.4	12.4	4.1	3.91
↓	5	0.3	1.9	19.3	5.9	0.2	1.4	12.2	3.7	3.89
1.37	6	0.1	1.1	19.1	5.7	0.1	0.8	12.1	3.6	3.88

Table II

Average Waiting Times and Total Delays (in Seconds) Using BISS Specification Model for Verification on Entry Only

(Continued)

		Entrance				Exit				
		Wait		Wait + Proc		Wait		Wait + Proc		
AF	NB	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg (sec)	STDEV	Avg Throughput (people/min)
1.66	2	41.5	88.3	59.4	88.5	9.7	18.0	20.4	17.9	4.19
	3	4.0	9.3	22.6	10.7	2.3	5.3	13.7	5.9	4.01
	4	1.3	4.3	20.1	7.0	0.8	2.8	12.6	4.3	3.93
	5	0.5	2.4	19.4	6.1	0.3	1.7	12.2	3.8	3.90
1.66	6	0.2	1.3	19.1	5.8	0.1	1.0	12.1	3.6	3.88

Table 12

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using TI-Modeled Verifications on Both Entry and Exit

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
0.37	1	66	96	88	119	70	106	91	129
	2	18	29	42	56	18	29	42	56
	3	6	15	33	43	6	15	33	43
	4	0	5	30	41	0	5	30	41
	5	0	0	29	41	0	0	29	41
0.37	6	0	0	29	41	0	0	29	41
0.52	1	94	139	115	159	104	158	124	178
	2	21	33	44	59	22	33	44	60
	3	10	16	34	44	10	16	34	44
	4	0	8	30	41	0	8	30	41
	5	0	0	29	41	0	0	29	41
0.52	6	0	0	29	40	0	0	29	40
0.8	1	384	402	555	573	539	740	556	758
	2	30	45	52	71	31	47	53	74
	3	13	19	36	46	13	19	36	46
	4	2	12	31	42	2	12	31	42
	5	0	2	29	40	0	2	29	40
0.8	6	0	0	29	40	0	0	29	40
0.94	2	36	54	58	81	40	62	62	89
	3	14	21	37	48	14	21	37	48
	4	5	12	31	42	5	12	32	42
	5	0	4	30	40	0	5	30	41
0.94	6	0	0	29	40	0	0	29	40
1.0	2	39	59	61	85	44	71	67	98
	3	15	22	38	49	15	23	38	50
	4	6	13	32	43	6	13	32	43
	5	0	5	30	41	0	6	30	41
1.0	6	0	0	29	41	0	0	29	41

Table 12

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using TI-Modeled Verifications on Both Entry and Exit
(Continued)

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
1.09	2	46	69	68	94	56	92	78	116
	3	16	24	39	50	16	25	39	51
	4	7	14	32	43	7	14	32	43
	5	0	7	30	41	0	7	30	41
	6	0	0	29	41	0	0	29	41
1.37	2	94	148	116	169	161	258	179	277
	3	21	32	43	59	23	36	45	64
	4	10	16	34	44	11	17	34	45
	5	3	9	31	41	3	9	31	42
1.37	6	0	3	29	41	0	3	29	41
1.66	2	332	467	351	486	525	698	544	717
	3	29	44	50	72	37	65	60	93
	4	13	20	36	47	13	21	37	48
	5	5	11	31	42	6	12	31	42
1.66	6	0	6	30	41	0	6	30	41

Table 13

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using BISS Specification Model for Verification on Both Entry and Exit

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
0.37	1	62	87	81	105	64	90	83	109
	2	19	29	40	49	19	29	40	49
	3	5	15	33	38	5	15	33	38
	4	0	5	31	35	0	5	31	35
	5	0	0	30	34	0	0	30	34
0.37	6	0	0	30	34	0	0	30	34
0.52	1	86	123	105	141	95	141	113	159
	2	21	33	42	52	22	33	42	53
	3	10	17	33	39	10	17	33	39
	4	0	9	31	36	0	9	31	36
	5	0	0	30	34	0	0	30	34
0.52	6	0	0	30	34	0	0	30	34
0.8	1	368	529	386	547	564	767	582	786
	2	29	42	48	61	31	45	50	64
	3	13	20	35	42	13	20	35	42
	4	3	12	31	36	3	12	31	36
	5	0	3	30	35	0	3	30	35
0.8	6	0	0	30	34	0	0	30	34
0.94	2	35	50	54	69	38	57	57	75
	3	15	22	36	43	15	22	36	43
	4	5	13	32	37	5	13	32	37
	5	0	5	30	35	0	5	30	35
0.94	6	0	0	30	35	0	0	30	35
1.0	2	38	54	56	72	43	66	61	84
	3	15	23	36	49	16	24	37	50
	4	6	13	32	37	6	14	32	37
	5	0	6	30	35	0	6	30	35
1.0	6	0	0	30	34	0	0	30	34
1.09	2	44	63	62	81	54	88	72	106
	3	16	24	37	45	17	25	38	46

Table 13

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using BISS Specification Model for Verification on Both Entry and Exit

(Continued)

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
	4	7	14	32	37	7	14	32	37
	5	0	7	30	35	0	7	30	35
1.09	6	0	0	30	35	0	0	30	35
1.37	2	92	145	111	164	166	262	184	281
	3	21	31	41	50	23	35	43	54
	4	10	16	33	39	11	17	33	39
	5	3	9	31	36	3	10	31	36
1.37	6	0	3	30	35	0	3	30	35
1.66	2	340	477	358	496	545	704	563	723
	3	29	42	48	61	37	62	55	81
	4	13	20	35	41	14	22	35	43
	5	6	12	31	36	6	12	32	37
1.66	6	0	6	30	35	0	6	30	35

Table 14

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using TI-Modeled Verifications on Entry Only

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
0.37	1	58	84	80	109	36	52	47	63
	2	17	26	40	52	11	18	24	30
	3	2	14	32	42	1	9	20	24
	4	0	2	30	41	0	1	19	22
	5	0	0	29	41	0	0	18	21
0.37	6	0	0	29	40	0	0	18	21
0.52	1	78	117	100	140	44	65	55	76
	2	19	29	42	56	13	20	25	31
	3	7	15	33	43	5	10	20	24
	4	0	6	30	41	0	4	19	22
	5	0	0	29	40	0	0	19	21
0.52	6	0	0	29	40	0	0	18	21
0.8	1	274	412	293	430	83	130	94	140
	2	25	38	47	65	16	24	28	35
	3	11	17	35	45	8	12	21	25
	4	0	9	31	41	0	7	19	22
	5	0	0	29	40	0	0	19	22
0.8	6	0	0	29	40	0	0	18	21
0.94	1	668	911	686	929	155	246	165	257
	2	29	43	51	71	18	26	29	38
	3	12	19	36	46	9	13	22	26
	4	2	11	31	42	1	8	19	22
	5	0	2	29	40	0	0	19	22
0.94	6	0	0	29	40	0	0	18	21
1.0	1	873	998	891	998	196	316	206	327
	2	32	48	54	76	19	28	30	39
	3	13	19	36	46	9	13	22	26
	4	3	11	31	42	2	8	19	23
	5	0	3	29	41	0	2	19	22
1.0	6	0	0	29	40	0	0	19	21

Table 14

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using TI-Modeled Verifications on Entry Only

(Continued)

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
1.09	1	998	999	998	999	312	475	323	486
	2	36	57	59	84	21	30	32	41
	3	14	20	37	47	9	14	22	26
	4	4	12	31	42	2	8	19	23
	5	0	4	29	41	0	2	19	22
1.09	6	0	0	29	41	0	0	19	21
1.37	2	70	117	93	140	27	41	38	51
	3	17	26	40	54	11	16	23	28
	4	7	14	32	43	5	9	20	23
	5	0	6	30	41	0	5	19	22
1.37	6	0	1	29	40	0	0	19	21
1.66	2	256	374	274	393	45	74	55	84
	3	22	35	45	64	13	19	25	31
	4	10	16	34	45	7	11	20	24
	5	3	9	31	42	2	7	19	22
1.66	6	0	3	29	41	0	2	19	22

Table 15

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using BISS Specification Model for Verification on Entry Only

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
0.37	1	56	79	75	97	34	49	45	60
	2	17	26	38	47	11	18	24	30
	3	3	15	32	38	32	38	20	23
	4	0	2	30	35	0	0	19	22
	5	0	0	30	34	0	0	19	21
0.37	6	0	0	30	34	0	0	18	21
0.52	1	73	103	91	122	43	60	53	71
	2	19	30	40	50	12	19	25	31
	3	7	16	33	38	5	10	20	24
	4	0	0	31	35	0	0	19	22
	5	0	0	30	34	0	0	19	21
0.52	6	0	0	30	34	0	0	18	21
0.8	1	255	403	273	421	77	118	87	128
	2	24	36	44	56	16	25	28	35
	3	11	17	34	40	8	12	21	25
	4	0	10	31	36	0	7	19	22
	5	0	1	30	34	0	0	19	22
0.8	6	0	0	30	34	0	0	18	21
0.94	1	651	851	669	869	150	239	160	249
	2	29	42	48	61	18	26	29	37
	3	13	19	35	41	8	12	21	25
	4	2	11	31	36	0	8	19	23
	5	0	0	30	35	0	0	19	22
0.94	6	0	0	30	34	0	0	19	21
1.0	1	868	998	887	999	201	322	211	333
	2	31	45	50	65	19	27	30	38
	3	13	20	35	42	9	13	22	26
	4	3	11	31	36	1	8	19	23
	5	0	3	30	35	0	2	19	22
1.0	6	0	0	30	34	0	0	18	21

Table 15

95th and 98th Percentiles (In Seconds) for Waiting Times and Total Delays Using BISS Specification Model for Verification on Entry Only

(Continued)

		Entrance				Exit			
		Wait		Wait + Proc		Wait		Wait + Proc	
AF	NB	95	98	95	98	95	98	95	98
1.09	1	998	999	998	999	342	496	353	507
	2	35	52	54	71	20	29	31	40
	3	14	21	35	42	9	14	22	26
	4	4	12	31	36	2	8	20	23
	5	0	4	30	35	0	2	19	22
	6	0	0	30	34	0	0	19	21
	2	62	103	81	121	27	40	38	50
	3	17	25	38	46	11	16	23	28
	4	7	14	32	37	5	10	20	23
	5	0	6	30	35	0	5	19	22
1.09	6	0	1	30	34	0	0	19	21
1.66	1	999	999	999	999	999	999	999	999
	2	243	359	262	378	42	65	52	76
	3	22	34	42	54	13	19	25	31
	4	10	16	33	39	7	11	20	24
	5	3	9	31	36	1	7	19	22
1.66	6	0	3	30	34	0	2	19	21

times is a direct consequence of a greater percentage of "short" transactions that occur when someone is waiting to enter a booth that someone is exiting.

Tables 16 and 17 give histograms of both waiting time alone and the processing time alone as a function of time. Both tables are for two booths, an AF of one, using the TI modeled verification. Table 16 is for verification both on entry and exit, and Table 17 is for verification on entry alone.

Table 16

OF BOOTHS = 2; ARR FACTR =1.00; VER ON EN = T; VER ON EX = T; MODEL = T

	ENTRANCE						EXIT			
	BEGIN RATES		WAITING		PROCESSING		WAITING		PROCESSING	
TIME PERIOD	ENTR	EXIT	AVE	STDEV	AVE	STDEV	AVE	STDEV	AVE	STDEV
6:00- 6:30	0.30	0.30	2.29	7.18	18.98	10.96	2.40	7.09	18.97	11.83
6:30- 7:00	2.15	0.30	8.48	15.85	18.79	10.18	4.82	9.60	17.30	10.26
7:00- 7:30	4.00	0.30	15.98	24.56	18.66	10.02	6.88	11.21	16.17	9.63
7:30- 8:00	4.00	0.50	11.67	22.06	18.58	10.43	5.73	11.08	17.32	10.50
8:00- 8:30	2.00	1.00	5.76	12.09	18.42	10.44	5.20	10.90	17.96	10.59
8:30- 9:00	1.50	1.00	4.80	11.00	18.33	10.24	4.42	10.72	18.42	10.86
9:00- 9:30	1.00	1.50	4.23	10.36	18.35	10.45	4.29	9.86	18.43	10.47
9:30-10:00	1.00	1.00	3.93	9.42	18.49	10.41	3.56	8.62	18.61	10.94
10:00-10:30	1.00	1.00	4.05	9.48	18.17	9.45	4.24	10.04	18.46	10.16
10:30-11:00	1.00	1.50	4.24	9.76	18.04	9.68	4.93	10.84	18.56	10.82
11:00-11:30	1.00	1.50	5.48	11.27	17.66	10.39	7.16	13.85	18.36	10.17
11:30-12:00	1.00	3.00	7.27	12.84	17.36	10.51	9.90	17.54	18.06	10.19
12:00-12:30	2.00	2.00	7.53	14.71	17.63	9.96	7.78	14.83	18.04	10.16
12:30-13:00	1.00	2.00	7.89	14.59	17.56	10.05	8.06	15.00	17.93	10.19
13:00-13:30	3.00	2.00	12.21	20.58	17.76	10.57	9.55	16.26	17.13	10.11
13:30-14:00	2.00	1.50	5.95	13.12	18.24	10.26	5.86	11.29	18.26	11.04
14:00-14:30	1.00	1.00	3.65	8.96	18.44	10.34	3.72	9.06	18.54	10.45
14:30-15:00	1.00	1.00	3.78	9.58	18.50	10.54	3.61	8.86	18.50	10.91
15:00-15:30	1.00	1.00	3.55	8.89	18.55	10.40	3.73	9.00	18.43	9.81
15:30-16:00	1.00	1.00	6.05	12.11	18.02	10.35	6.58	12.96	18.08	10.56
16:00-16:30	2.00	3.00	11.52	17.52	16.26	10.88	22.23	30.78	17.84	10.11
16:30-17:00	1.00	5.00	7.00	11.58	16.49	9.35	20.36	35.26	18.55	10.33
17:00-17:30	0.30	2.00	3.06	7.88	18.03	10.18	3.79	9.54	18.89	10.30
17:30-18:00	0.30	1.15	1.89	6.29	18.81	10.37	2.32	7.31	18.85	9.95

<u>**Table 17**</u>

OF BOOTHS = 2; AMP FACTR =1.00; VER ON EN = T; VER ON EX = F; MODEL = 1

	ENTRANCE						EXIT			
	BEGIN RATES		WAITING		PROCESSING		WAITING		PROCESSING	
TIME PERIOD	ENTR	EXIT	AVE	STDEV	AVE	STDEV	AVE	STDEV	AVE	STDEV
6:00- 6:30	0.30	0.30	2.15	6.47	19.05	10.73	1.42	4.39	11.78	3.59
6:30- 7:00	2.15	0.30	7.90	14.72	18.91	10.52	3.92	7.57	10.31	3.71
7:00- 7:30	4.00	0.30	14.78	22.59	18.63	9.97	5.60	8.59	9.45	3.66
7:30- 8:00	4.00	0.50	8.83	17.14	18.60	10.08	4.17	7.84	10.53	3.76
8:00- 8:30	2.00	1.00	4.33	9.67	18.60	10.18	3.06	6.86	11.15	3.76
8:30- 9:00	1.50	1.00	3.62	8.62	18.49	9.78	2.75	6.49	11.43	3.72
9:00- 9:30	1.00	1.50	2.83	7.61	18.52	9.69	2.23	5.59	11.62	3.68
9:30-10:00	1.00	1.00	2.80	7.36	18.66	10.52	2.19	5.69	11.61	3.67
10:00-10:30	1.00	1.00	3.08	8.20	18.66	10.41	2.43	5.97	11.56	3.65
10:30-11:00	1.00	1.50	3.10	7.81	18.70	11.70	2.64	6.25	11.61	3.71
11:00-11:30	1.00	1.50	3.38	7.99	18.33	10.30	3.26	7.56	11.53	3.65
11:30-12:00	1.00	3.00	4.53	9.76	17.94	10.20	4.12	8.31	11.38	3.76
12:00-12:30	2.00	2.00	4.24	9.28	18.16	10.70	3.71	7.69	11.31	3.72
12:30-13:00	1.00	2.00	5.23	10.55	18.03	9.96	4.05	8.20	11.19	3.79
13:00-13:30	3.00	2.00	7.51	14.15	18.05	10.39	5.33	9.66	10.62	3.86
13:30-14:00	2.00	1.50	3.78	8.84	18.38	9.47	2.90	6.60	11.29	3.71
14:00-14:30	1.00	1.00	2.73	7.41	18.66	9.74	2.16	5.67	11.61	3.69
14:30-15:00	1.00	1.00	2.78	7.61	18.67	10.64	2.19	5.72	11.58	3.66
15:00-15:30	1.00	1.00	2.99	8.04	18.71	10.32	2.22	5.76	11.63	3.70
15:30-16:00	1.00	1.00	4.19	9.72	18.27	10.26	3.47	7.40	11.30	3.75
16:00-16:30	2.00	3.00	5.36	9.80	17.32	10.41	6.38	10.82	11.27	3.76
16:30-17:00	1.00	5.00	3.75	8.48	17.67	10.51	4.32	8.96	11.74	3.68
17:00-17:30	0.30	2.00	1.83	5.93	18.65	10.49	1.76	4.99	11.89	3.56
17:30-18:00	0.30	1.15	1.65	5.79	18.95	11.22	1.21	4.08	11.93	3.61

Appendix II

PROMPTING WORD SELECTION TRADE-OFF STUDY

Appendix II

PROMPTING WORD SELECTION TRADE-OFF STUDY

The results presented in this section are based upon experimental data extracted from the following six sources:

1. Phase I analysis at Texas Instruments (July 1974)
2. MITRE Phase I evaluation reference files (Oct-Nov 1975)
3. MITRE Phase II evaluation (Aug-Oct 1976)
4. 21 session data set from 11 speakers (1979)
5. CIC entry control system references and casual impostor sessions from a subset of CIC users (1979)
6. CIC entry control system reference files (primarily 1979)

For conciseness, the sources will be referred to by number throughout this section. These sources of data represent the three different sets of prompting words given in Table 1. Source 1 used the words in set (A); sources 2-3 used the words in set (B); sources 4-6 used the words in set (C).

TABLE 1
SETS OF PROMPTING WORDS

SET A.	COOL	BIRDS	STOPPED	WEST
	SMALL	BUGS	SING	DOWN
	HUGE	TWIGS	SANG	DEEP
	STRANGE	TOADS	STOOD	WILD
SET B.	NORTH	LAWN	GREAT	CAMP
	SOUTH	LIMB	WIDE	POINT
	EAST	RUN	GOOD	PLOT
	FIRST	ROOM	WEST	CUBE
SET C.	GOOD	BRUCE	CALLED	HARD
	PROUD	BEN	SWAM	NEAR
	STRONG	JOYCE	CAME	NORTH
	YOUNG	JEAN	SERVED	HIGH

The types of data readily available during this evaluation were not the same from all six sources. Useful information would include the following:

1. Actual reference patterns for all enrolled speakers
2. Squared distances (scanning errors) for each word for each true speaker verification session (TYPE I trial)
3. Squared distances (scanning errors) for each word for each impostor session (TYPE II trial)
4. Number of verification sessions
5. Average EHAT's (expected scanning errors) for each word for each speaker
6. Equal error rate (EER) for each word for each sex and type of preprocessing.

To clarify which of the above data are available from each of the six

sources, a short description of each source is given below. Note that all impostor trials used only reference files for speakers of the same sex.

I. DATA SOURCE DESCRIPTIONS

1. The phase 1 analysis at TI was based on 63 male references. Although 73 male speakers were enrolled, only 63 had verification sessions. The only available information from this data set that pertained to this study were equal error rates, average scanning errors, and number of sessions for each speaker. Every session consisted of four phrases (each word used once for each session). The true speaker scanning errors used in calculating both the average scanning errors and the equal error rates were from all sessions except post-enrollment (the first four sessions after enrollment) for all of the 63 true speakers. The impostor scanning errors used in determining equal error rates were found by comparing two sessions from 60 of the 63 speakers to the reference files at the end of the experiment (after all adaptation) for the other 62 speakers, for a total of 7440 impostor trials for each word. The number of true speaker sessions from which the reference files were derived ranged from 8 to 130, with an average of 25.5 sessions per speaker.

2. The data available from the phase I MITRE evaluation consisted of listings of the reference files at the completion of the test. Although the actual reference patterns existed in documentation, they were not in a machine readable form. The only information extracted from these listings consisted of the average EHAT's for each word for each speaker, the sex of each speaker, and the number of sessions for each speaker. These data are from 209 speakers (170 male, 39 female). The average number of sessions was 14.1 for males and 12.2 for females. The average number of phrases across all speakers was 4.77 during post-enrollment and was 1.70 phrases for subsequent processing.

3. More extensive data for the phase II MITRE evaluation existed than for phase I. Subsets of both the impostor and true speaker trials were punched on cards from listings, which included other information such as EHAT's, phrase number and sesison number. The number of enrolled speakers was 199 (164 males, 35 females). The number of impostors was 78 (71 males, 7 females) from the set of enrolled speakers, which were compared to 112 (106 males, 6 females) reference patterns. The number of impostor trails available, however, was not the N*(N-1) possible, but was a subset as shown in Table 2. The actual numbers of impostor and true speaker trials is shown in Table 3.

TABLE 3
NUMBER OF IMPOSTOR AND TRUE SPEAKER TRIALS USED
FROM MITRE PHASE II EXPERIMENTS

	SESSIONS		PHRASES		AVE NO. PER WORD	
	TOTAL	NON PE	TOTAL	NON PE	TOTAL	NON PE
MALE IMPOSTORS	713	—	3366	—	842	—
MALE TRUE SPEAKERS	2373	1781	8262	5962		1466
FEMALE IMPOSTORS	13	—	69	—	17	—
FEMALE TRUE SPEAKERS	498	372	1954	1408		352

4. The fourth data set was collected from 11 speakers (6 males, 5 females)

in the sound booth in the TI Speech Research Lab. Each speaker contributed 21 sessions after enrollment. The primary purpose of these data was to provide a set of actual speech data that could be massaged through different filters, preprocessings, etc. However the limited number of speakers made the general applicability of the results somewhat questionable.

5. The fifth data set was collected using the TI entry control booth. This data set included two sets of specially collected impostor data: one set of 40 males, using BISS-type preprocessing, and one set of 23 males and 9 females using IPMOD2-type preprocessing. Each impostor was enrolled on the system when the impostor data were collected, but were not necessarily in the reference data set. Each impostor had one session with three repetitions of four phrases, each of the 16 words occurring once in the four phrase set. The BISS-type preprocessed impostors were compared against reference files for 258 male speakers enrolled on the CIC system as of September 1978. For IPMOD2, the 23 male impostors were compared against 71 male references and the 9 female impostors were compared against 14 female references, where the references were as of 3 January, 1980 and the true speaker data were from the CIC booth usage for the previous 6 weeks.

6. The final data set used was just the CIC reference files as of 25 October 1979. The only purpose of this was to calculate various attributes of the reference files for the speakers that might have shown some correlation with the equal error rates determined for the CIC data in the prior experiment. These reference files as of 25 October 1979 contained references for 193 speakers (170 males, 23 females) using BISS-type preprocessing and 74 speakers (61 males, 13 females) using IPMOD2-type preprocessing.

II. ANALYSIS

Shown in Figure 1 is the general form for the true speaker (TYPE I) and impostor (TYPE II) error rates as a function of distance between an imput and a reference. The task then is to find an easily measurable parameter which has a well defined relationship to the error rate so that prompting words with low equal error rates can be chosen without performing a large scale experiment with many true speaker and impostor trials. Since all six of the data bases discussed above have the average true speaker distances available (or at least the EHAT's), that was the first parameter tried. Tables of the normalized average squared distances are given in Tables 4-7 and EERs are given in Tables 8-10. These are plotted in Figure 2. (A "*" indicates those plots with sample sizes judged to be adequate.) No distinct relationship between the two variables can be seen. Considering Figure 1, this is an expected disappointment since in this case we are trying to find a relationship between the average of a distribution and the overlap of its tail with another distribution.

The same problem exists using the "median" of the true speaker distribution, as shown in Figure 3, a plot of EER vs the medians of the true speaker squared distances for experiment 5. Figures 4 and 5 show plots of EER vs equal error threshold and 10th percentile of the true speaker squared distances respectively. Although both of these distribution parameters are more closely associated with the tails of the distributions, the only improvements are slightly more skewed plots in Figure 4. The actual true speaker and impostor distributions from which these data were derived are shown in Figures 6-8.

The data used to plot Figures 3-5 are shown in tabular form in Tables 11-13. In addition to the true speaker distribution parameters and the equal error thresholds, these tables also contain three additional columns showing data associated with the impostor distributions. The first of these columns is the 10th percentile values of the squared distances for the impostor distributions for experiment 5. The last two columns were derived from data in experiment 6. Although the reference files are not identical between experiments 5 and 6, the supposition is made that the statistics of the reference files are similar since many of the enrolled speakers are the same and the number of speakers is non-negligible. The next to the last column represents the 30th percentile of the squared distances between the reference files for each word and the average reference file for the corresponding word across all speakers. The final column gives the 10th percentile of the squared differences between the reference file for each speaker for a given word and the reference file for every other speaker for the same word. The EER obtained from experiment 5 is plotted in Figures 9-11 vs. the values given in these last three columns of Tables 11-13. Although the data in these plots are still scattered, an inverse relationship between the EER and the distance between reference files is much more evident than in the prior data. This suggests that one mechanism that could be used in selecting good prompting words is to collect single session data from a large populaton and select those words having the largest value at say the 10th percentile of the distances computed between all pairs of patterns for each word. (Note that although not presented here, the averages of these distances did not show such a nice relationship.)

A final approach tried was to try to find some inherent property of the patterns for the words which related to the equal error rates. Tables 14-16 show the relationship between the EERs derived from experiment 5 and the following four measures:

1. Place of articulation of the vowel (high/low/medial and front/back/center)
2. Percentage (P(I)) of each quantized energy level, I (I=0,...,7), across all filters
3. A measure of the average density across all reference patterns for a given word (I*P(I))
4. A measure of the squared deviation from the average energy level (P(I) times the square of (1 + the integer part of the magnitude of (3.5-I))).

This last measure was prompted by the idea that since we are using a squared distance measure in the verification algorithm, a bland pattern that had many filters with medial values would, on the average, have lower values of squared distances for impostors. As can be seen from the tables, none of these measures seem particularly valuable.

One final approach was to try to relate the EERs to the values of the formant locations for the vowels in the patterns. This is done in Table 17, which gives the center frequencies of the 14 filters and the center frequencies of the formants of the vowels in the central section of each of the prompting words. (Two lines are used in the table for initial and final locations for diphthongs.) Immediately obvious is the generally increasing EERs as the values of the second and third formants increase. This may be due to the lack of proper amplitude compensation for the wider bandwidth top three filters in the preprocessing algorithm. The glaring exception to this trend in the second formant

is for the words containing the vowel "AW" (called, strong). The problem in these cases becomes obvious by investigating some actual reference patterns for these two words. The gentle roll-off of the filters causes a smearing of the energies out of filters 2-4 such that an impostor does not encounter such drastic energy differences as the locations of the first and second formants vary, resulting in low scanning errors for these words, even for impostors.

One final observation can be made from the EERs in the *'ed columns of Tables 8-10. This observation is that, for the same vowel, those words having a prevocalic "R" tend to have lower relative EERs than those without the formant movement caused by the adjacent "R".

III. CONCLUSIONS

Based upon the above sets of experiments, the following guidelines should be used in the selection of sets of monosyllabic prompting words for voice authentication systems:

1. One session of all candidate words should be collected from several hundred speaker and those words having low relative between speaker squared distances at say the 10th percentile should be excluded.
2. For the BISS* filter bank, words should be excluded whose central vowel has either of the following properties:
   A. Third formant above 2700 Hz and second formant above 2000 Hz (EI as in "CAME" and EE as in "JEAN")
   B. Close first and second formants throughtout the duration of the vowel (AW as in "CALLED" or "STRONG").
3. Words having prevocalic "R's" should be favored over those without.
4. Words having little preceptual difference should not both be used (such as "RUN" and "ROOM"), since it has been observed from listening to a small sample of the tapes of true speaker trials from the MITRE phase II experiments, that speakers sometimes confuse such pair of words.

In addition to these rules, the adjective-noun-verb-adverb paradigm was found to be preferable to the adjective-noun-adjective-noun paradigm used on BISS, also concluded from the MITRE tapes, which showed a non-negligible number of trials in which the subject paused so long between the middle two words that the system thought the speaker had finished speaking.

One additional rule in choosing word sets is prompted by the reasonable sounding conjecture that since initial reference point determination during enrollment is made based upon the location of energy peaks, the elimination of between word nasal/semivowel/glide bridges would minimize the energy-based, syllable detection problem during enrollment.

---

* The "BISS" FILTER BANK is as defined in Table I, page 20, of "SPEAKER VERIFICATION," RADC-TR-74-179, April 1974. These filters have constant bandwidths of 220 Hz and constant spacing between center frequencies of 175 Hz. Note that the top two filters (C.F.'s OF 2805 AND 2980 Hz) are not used in either the BISS-type or the IPMOD2-type preprocessing.

Finally, due to the importance of this type of investigation in terms of the potential for reducing error rates, this summary should be considered only an interim report in a continuing investigation. Some of the tasks in this continuing investigation are as follows:

1. See if EERs for the words vary with the choice of center frequencies and bandwidths of the filter bank.
2. Choose some substitute words for
   A. Jean (Jan, June, Joan)
   B. Called
   C. Strong
   D. Came.
3. Consider word set modifications that eliminate the nasal bridge that can exist between the third and fourth words of word set "C".
4. Consider using some two-syllable prompting words
   A. Advantage - Performance improves as the amount of speech data increases
   B. Disadvantage - Enrollment algorithm that looks for four energy peaks would have to be modified.
5. After choosing several candidate substitute words, collect data from the approximately 40 males at the TI Hillcrest location and perform the between speaker distance measure experiment proposed at the beginning of this section.
6. Consider methods for quantitatively measuring the perceptual difference between words in a word set.
7. Consider allowing present tense verbs (either partially or totally) in place of the past tense verbs currently used.

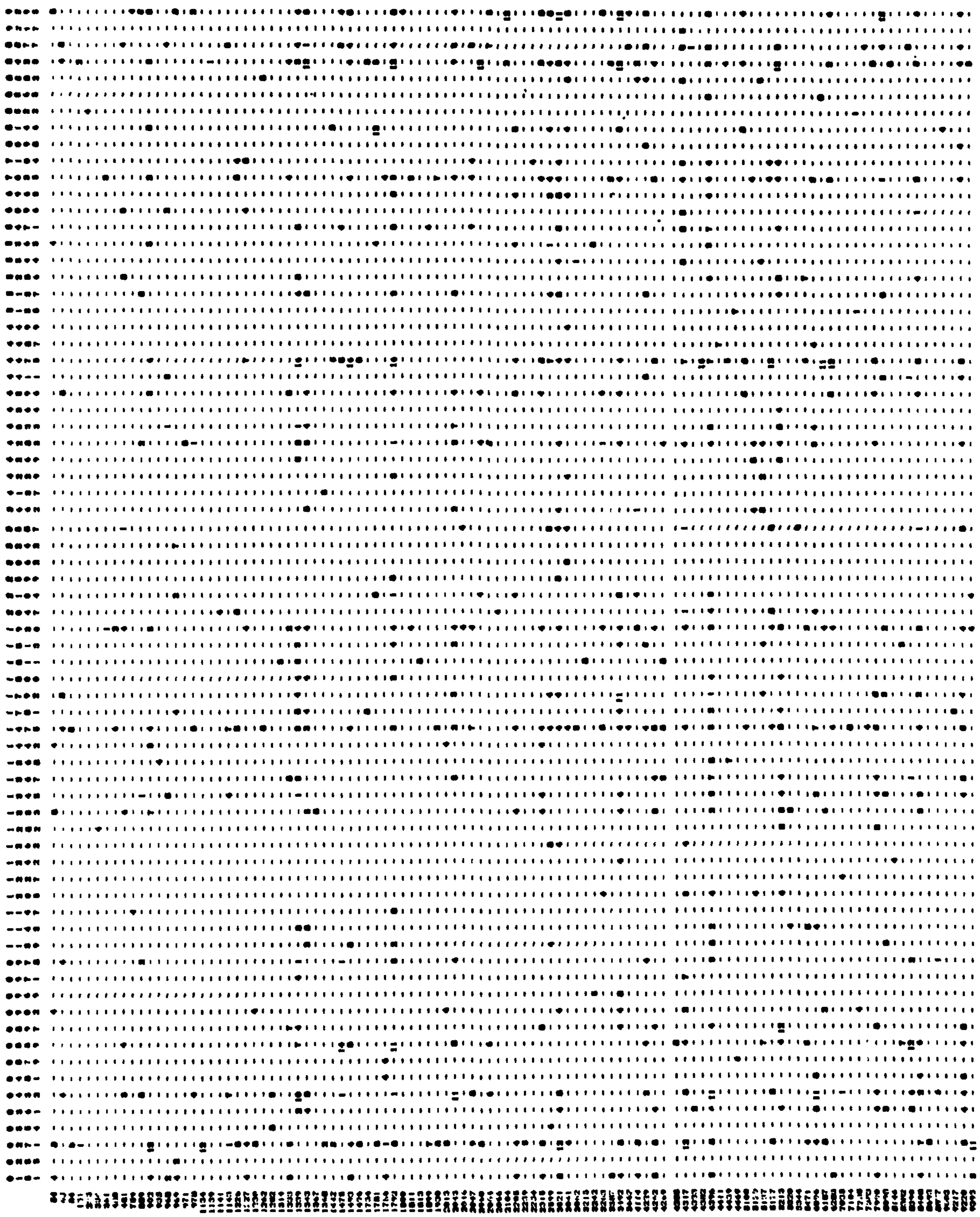


Table II(b)

	4406	5267	6723	7350	8094	8697	9343
6270	-	-	8	-	-	-	-
7141	-	-	4	8	3	8	-
8024	-	-	8	-	-	-	4
8359	-	-	-	8	-	-	-
8441	2	-	4	4	4	-	-
8697	-	4	-	-	-	-	-

7 IMPOSTORS COMPARED TO 6 TRUE SPEAKERS OUT OF 36 SPKRS OF THIS SEX ENROLLED

TABLE 4. NORMALIZED AVERAGE SQUARED DISTANCE FOR TRUE SPEAKERS FROM THEIR REFERENCES FOR MALES WITH BISS-TYPE PREPROCESSING

	NORMALIZED AVERAGE TRUE SPEAKER DISTANCES FOR DATA BASE X / WORD SET Y					WORDS		
VOWEL	1/A	*2/B	4/C	5/C	*6/C	A	B	C
EE	.78	.87	.89	.71	.74	DEEP	EAST	JEAN
IX	.90	.86	.74	.65	.65	TWIGS	LIMB	NEAR
	1.00					SING		
EH	.84	.74	.82	.79	.81	WEST	WEST	BEN
AE	.79	.86	1.00	1.00	1.00	SANG	CAMP	SWAM
AA	.67	.41	.60	.47	.50	STOPPED	PLOT	HARD
AW	.35	.46	.37	.40	.42	SMALL	LAWN	CALLED
			.81	.61	.59			STRONG
UX	.99	1.00	.95	.92	.96	STOOD	GOOD	GOOD
UU	1.00	.90	.69	.77	.84	COOL	ROOM	BRUCE
UH	.70	.77	.96	.91	.92	BUGS	RUN	YOUNG
IR	.78	.67	.67	.69	.69	BIRDS	FIRST	SERVED
AU	.73	.60	.86	.80	.83	DOWN	SOUTH	PROUD
AI	.61	.66	.49	.56	.55	WILD	WIDE	HIGH
OI		.67	.88	.86	.86		POINT	JOYCE
IU	.87	.82				HUGE	CUBE	
OU	.79					TOADS		
EI	.86	.77	.91	.73	.77	STRANGE	GREAT	CAME
OX		.71	.75	.63	.65		NORTH	NORTH
MAX.	125	164	118	139	125			
	AVE.		91.4	99.9	92.0			

TABLE 5. NORMALIZED AVERAGE SQUARED DISTANCE FOR TRUE SPEAKERS FROM THEIR REFERENCES FOR FEMALES WITH BISS-TYPE PREPROCESSING

	NORMALIZED AVERAGE TRUE SPEAKER DISTANCES FOR DATA BASE X / WORD SET Y				WORDS	
VOWEL	*2/B	4/C	5/C	*6/C	B	C
EE	.87	1.00	.94	.94	EAST	JEAN
IX	.85	.82	.91	.84	LIMB	NEAR
EH	.77	.99	.93	.94	WEST	BEN
AE	.88	.81	.95	.94	CAMP	SWAM
AA	.40	.46	.41	.39	PLOT	HARD
AW	.46	.31	.34	.35	LAWN	CALLED
		.71	.82	.85		STRONG
UX	1.00	.65	.90	.83	GOOD	GOOD
UU	.89	.65	.78	.81	ROOM	BRUCE
UH	.75	.95	1.00	1.00	RUN	YOUNG
IR	.62	.59	.74	.78	FIRST	SERVED
AU	.54	.85	.83	.83	SOUTH	PROUD
AI	.67	.74	.60	.58	WIDE	HIGH
OI	.65	.82	.81	.85	POINT	JOYCE
IU	.85				CUBE	
EI	.77	.90	.81	.83	GREAT	CAME
OX	.64	.65	.54	.59	NORTH	NORTH
MAX.	165	141	152	142		
		104.9	116.9	109.6		

* EHATS (EXPECTED SCANNING ERRORS)

TABLE 6. NORMALIZED AVERAGE SQUARED DISTANCE FOR TRUE SPEAKERS FROM THEIR REFERENCES FOR MALES WITH IPMOD2-TYPE PREPROCESSING

	NORMALIZED AVERAGE TRUE SPEAKER DISTANCES FOR DATA BASE X / WORD SET Y				WORDS	
VOWEL	3/B	4/C	5/C	*6/C	B	C
EE	.81	.81	.79	.71	EAST	JEAN
IX	.90	.72	.64	.60	LIMB	NEAR
EH	.67	.77	.73	.77	WEST	BEN
AE	1.00	1.00	1.00	1.00	CAMP	SWAM
AA	.45	.59	.45	.44	PLOT	HARD
AW	.43	.34	.40	.38	LAWN	CALLED
		.80	.58	.53		STRONG
UX	.72	.79	.96	.95	GOOD	GOOD
UU	.82	.55	.78	.73	ROOM	BRUCE
UH	.87	.89	.92	.90	RUN	YOUNG
IR	.57	.59	.69	.61	FIRST	SERVED
AU	.65	.84	.83	.78	SOUTH	PROUD
AI	.66	.51	.51	.53	WIDE	HIGH
OI	.68	.81	.83	.79	POINT	JOYCE
IU	.66				CUBE	
EI	.65	.93	.87	.79	GREAT	CAME
OX	.71	.68	.59	.59	NORTH	NORTH
MAX.	153	150	173	184		
		108.9	125.1	127.7		

TABLE 7. NORMALIZED AVERAGE SQUARED DISTANCE FOR TRUE SPEAKERS FROM THEIR REFERENCES FOR FEMALES WITH IPMOD2-TYPE PREPROCESSING

	NORMALIZED AVERAGE TRUE SPEAKER DISTANCES FOR DATA BASE X / WORD SET Y				WORDS	
VOWEL	3/B	4/C	5/C	*6/C	B	C
EE	.39	.78	.94	.85	EAST	JEAN
IX	.78	.66	.79	.81	LIMB	NEAR
EH	.62	.85	.87	.86	WEST	BEN
AE	.61	.81	1.00	.98	CAMP	SWAM
AA	.34	.42	.34	.34	PLOT	HARD
AW	.44	.27	.30	.28	LAWN	CALLED
		.63	.68	.73		STRONG
UX	.47	.52	.77	.76	GOOD	GOOD
UU	.74	.44	.67	.71	ROOM	BRUCE
UH	1.00	.87	.95	1.00	RUN	YOUNG
IR	.54	.48	.64	.69	FIRST	SERVED
AU	.67	1.00	.82	.81	SOUTH	PROUD
AI	.51	.77	.76	.61	WIDE	HIGH
OI	.56	.68	.76	.81	POINT	JOYCE
IU	.42	.49			CUBE	
EI	.57	.86	.72	.77	GREAT	CAME
OX	.75	.62	.49	.54	NORTH	NORTH
MAX.	241	189	204	218		
		133.2	146.6	157.4		

---

* EHATS (EXPECTED SCANNING ERRORS)

TABLE 8. EQUAL ERROR RATE AS FUNCTION OF WORD FOR SEVERAL DATA BASES FOR MALES & FEMALES WITH BISS-TYPE PREPROCESSING

	EQUAL ERROR RATE FOR DATA BASE X / WORD SET Y				WORDS	
VOWEL	*1/A/M	4/C/M	4/C/F	*5/C/M	A	C
EE	.225	.07	.12	.171	DEEP	JEAN
IX	.150	.06	.07	.160	TWIGS	NEAR
	.175				SING	
EH	.125	.05	.065	.136	WEST	BEN
AE	.165	.08	.09	.133	SANG	SWAM
AA	.140	.06	.185	.141	STOPPED	HARD
AW	.215	.10	.31	.190	SMALL	CALLED
		.07	.13	.196		STRONG
UX	.175	.05	.11	.116	STOOD	GOOD
UU	.180	.035	.115	.105	COOL	BRUCE
UH	.160	.09	.045	.121	BUGS	YOUNG
IR	.185	.065	.08	.137	BIRDS	SERVED
AU	.110	.085	.135	.118	DOWN	PROUD
AI	.185	.07	.135	.140	WILD	HIGH
OI		.05	.09	.116		JOYCE
IU	.165					HUGE
OU	.130					TOADS
EI	.115	.055	.08	.161	STRANGE	CAME
OX		.055	.11	.123		NORTH

TABLE 9. EQUAL ERROR RATE AS FUNCTION OF WORD FOR SEVERAL DATA BASES FOR MALES WITH IPMOD2-TYPE PREPROCESSING

	EQUAL ERROR RATE FOR DATA BASE X /WORD SET Y			WORDS	
VOWEL	*3/B	4/C	*5/C	B	C
EE	.318	.053	.173	EAST	JEAN
IX	.297	.063	.135	LIMB	NEAR
EH	.244	.045	.098	WEST	BEN
AE	.205	.086	.104	CAMP	SWAM
AA	.242	.050	.114	PLOT	HARD
AW	.269	.083	.160	LAWN	CALLED
		.053	.188		STRONG
UX	.285	.054	.088	GOOD	GOOD
UU	.254	.022	.089	ROOM	BRUCE
UH	.247	.081	.128	RUN	YOUNG
IR	.189	.067	.084	FIRST	SERVED
AU	.230	.075	.091	SOUTH	PROUD
AI	.246	.072	.121	WIDE	HIGH
OI	.286	.049	.103	POINT	JOYCE
IU	.285	.336		CUBE	
EI	.256	.042	.188	GREAT	CAME
OX	.233	.063	.076	NORTH	NORTH

* ADEQUATE (OR AT LEAST NOT INADEQUATE) SAMPLE SIZE

TABLE 10. EQUAL ERROR RATE AS FUNCTION OF WORD FOR SEVERAL DATA BASES FOR FEMALES WITH IPMOD2-TYPE PREPROCESSING

VOWEL	EQUAL ERROR RATE FOR DATA BASE X /WORD SET Y: *3/B	4/C	*5/C	WORDS: B	C
EE	.389	.111	.179	EAST	JEAN
IX	.353	.075	.123	LIMB	NEAR
EH	.296	.046	.165	WEST	BEN
AE	.176	.057	.212	CAMP	SWAM
AA	.235	.163	.149	PLOT	HARD
AW	.294	.250	.192	LAWN	CALLED
		.129	.203		STRONG
UX	.353	.100	.159	GOOD	GOOD
UU	.471	.125	.078	ROOM	BRUCE
UH	.309	.042	.099	RUN	YOUNG
IR	.294	.088	.102	FIRST	SERVED
AU	.278	.119	.134	SOUTH	PROUD
AI	.188	.144	.162	WIDE	HIGH
OI	.188	.055	.108	POINT	JOYCE
IU	.276	.278		CUBE	
EI	.278	.053	.206	GREAT	CAME
OX	.415	.065	.132	NORTH	NORTH

TABLE 11. TYPE I/TYPE II DISTRIBUTION PARAMETERS FOR DATA BASE 5 FOR MALES WITH BISS-TYPE PREPROCESSING

VOWEL	EER	TRUE SPEAKER DIST. AVE	MEDIAN	10%TILE	EQUAL ERROR THRESH	IMP. DIST. 10%TILE	REF FILE DIST. FROM AVE 30%TILE	AVE. REF FILE DIST. 10%TILE	WORD
UU	.105	.77	.80	.77	.83	.89	.83	.88	BRUCE
UX	.116	.92	.94	.92	.97	1.00	.77	.74	GOOD
OI	.116	.86	.86	.84	.88	.92	1.00	1.00	JOYCE
AU	.118	.80	.79	.83	.86	.89	.81	.81	PROUD
UH	.121	.91	.93	.88	.91	.94	.66	.69	YOUNG
OX	.123	.63	.61	.64	.67	.67	.70	.70	NORTH
AE	.133	1.00	1.00	1.00	1.00	.99	.95	.95	SWAM
EH	.136	.79	.76	.83	.81	.79	.66	.62	BEN
IR	.137	.69	.68	.73	.72	.69	.53	.54	SERVED
AI	.140	.56	.51	.61	.58	.54	.73	.56	HIGH
AA	.141	.47	.45	.50	.48	.46	.43	.40	HARD
IX	.161	.65	.61	.69	.63	.57	.74	.67	NEAR
EI	.161	.73	.70	.77	.72	.66	.60	.52	CAME
EE	.171	.71	.69	.76	.68	.61	.70	.57	JEAN
AW	.190	.40	.38	.44	.37	.30	.32	.30	CALLED
AW	.196	.61	.57	.64	.55	.46	.34	.39	STRONG
MAX.		139	122	225	206	189	151	218	

* ADEQUATE (OR AT LEAST NOT INADEQUATE) SAMPLE SIZE

TABLE 12. TYPE I/TYPE II DISTRIBUTION PARAMETERS FOR DATA BASE 5 FOR MALES WITH IPMOD2-TYPE PREPROCESSING

VOWEL	EER	TRUE SPEAKER DIST. AVE	TRUE SPEAKER DIST. MEDIAN	TRUE SPEAKER DIST. 10%TILE	EQUAL ERROR THRESH	IMP. DIST. 10%TILE	REF FILE DIST. FROM AVE 30%TILE	AVE. REF FILE DIST. 10%TILE	WORD
OX	.076	.59	.58	.60	.66	.71	.75	.75	NORTH
IR	.084	.69	.70	.69	.75	.77	.60	.61	SERVED
UX	.088	.96	.97	.95	1.00	1.00	.86	.79	GOOD
UU	.089	.78	.80	.77	.81	.80	.98	1.00	BRUCE
AU	.091	.83	.82	.90	.95	.94	1.00	.95	PROUD
EH	.098	.73	.69	.83	.86	.83	.77	.67	BEN
OI	.103	.83	.84	.83	.84	.78	.91	.88	JOYCE
AE	.104	1.00	1.00	1.00	1.00	.93	.99	1.00	SWAM
AA	.114	.45	.42	.52	.51	.43	.46	.47	HARD
AI	.121	.51	.45	.60	.58	.48	.67	.56	HIGH
UH	.128	.92	.93	.93	.89	.74	.68	.71	YOUNG
IX	.135	.64	.60	.69	.63	.51	.77	.66	NEAR
AW	.160	.40	.36	.45	.38	.27	.35	.28	CALLED
EE	.173	.79	.75	.84	.72	.51	.56	.53	JEAN
AW	.188	.58	.55	.63	.54	.36	.34	.37	STRONG
EI	.188	.87	.86	.94	.78	.55	.58	.55	CAME
MAX.		173	154	275	269	286	173	253	

TABLE 13. TYPE I/TYPE II DISTRIBUTION PARAMETERS FOR DATA BASE 5 FOR FEMALES WITH IPMOD2-TYPE PREPROCESSING

VOWEL	EER	TRUE SPEAKER DIST. AVE	TRUE SPEAKER DIST. MEDIAN	TRUE SPEAKER DIST. 10%TILE	EQUAL ERROR THRESH	IMP. DIST. 10%TILE	REF FILE DIST. FROM AVE 30%TILE	AVE. REF FILE DIST. 10%TILE	WORD
UU	.078	.67	.69	.66	.73	.77	.65	.58	BRUCE
UH	.099	.95	.97	.95	1.00	1.00	.88	.77	YOUNG
IR	.102	.64	.60	.69	.72	.70	.42	.46	SERVED
OI	.108	.76	.74	.77	.79	.76	1.00	.92	JOYCE
IX	.123	.79	.77	.83	.83	.80	.78	.58	NEAR
OX	.132	.49	.46	.56	.54	.45	.65	.75	NORTH
AU	.134	.82	.80	.85	.83	.75	.70	.59	PROUD
AA	.149	.34	.29	.40	.37	.21	.37	.24	HARD
UX	.159	.77	.74	.84	.76	.67	.37	.44	GOOD
AI	.162	.76	.68	.92	.81	.66	.66	.47	HIGH
EH	.165	.87	.85	.92	.86	.67	.97	1.00	BEN
EE	.179	.94	.96	.92	.84	.72	54	.54	JEAN
AW	.192	.30	.24	.39	.29	.19	.29	.18	CALLED
AW	.203	.68	.60	.85	.66	.45	.53	.54	STRONG
EI	.206	.72	.64	.88	.68	.32	.65	.81	CAME
AE	.212	1.00	1.00	1.00	.88	.67	.95	.80	SWAM
MAX.		204	189	315	299	303	172	295	

TABLE 14. PARAMETERS DERIVED FROM REFERENCE PATTERNS FOR MALES WITH BIGG-TYPE PREPROCESSING FROM DATA BASE 6.

VOWL	TYP	EER	% 0	% 1	% 2	% 3	% 4	% 5	% 6	% 7	I*P(I)	**
UU	HB	.105	0.30	0.14	0.11	0.10	0.11	0.10	0.08	0.06	2.45	8.75
UX	HB	.116	0.29	0.19	0.12	0.10	0.10	0.08	0.08	0.05	2.32	8.79
OI	MB/HF	.116	0.28	0.15	0.12	0.11	0.12	0.10	0.09	0.04	2.49	8.23
AU	LB/HB	.118	0.29	0.14	0.11	0.11	0.09	0.09	0.12	0.05	2.57	8.73
UH	MC	.121	0.27	0.19	0.13	0.11	0.10	0.09	0.08	0.04	2.37	8.41
OX	MB	.123	0.28	0.10	0.11	0.17	0.13	0.08	0.08	0.06	2.60	8.03
AE	LF	.133	0.22	0.20	0.15	0.13	0.11	0.08	0.07	0.03	2.37	7.58
EH	MF	.136	0.30	0.21	0.13	0.10	0.08	0.05	0.06	0.06	2.16	9.10
IR	HC	.137	0.30	0.14	0.11	0.10	0.10	0.10	0.08	0.06	2.47	8.92
AI	LB/HF	.140	0.36	0.10	0.07	0.06	0.09	0.09	0.15	0.08	2.67	10.11
AA	LB	.141	0.34	0.08	0.09	0.10	0.12	0.10	0.10	0.09	2.73	9.36
IX	HF	.160	0.31	0.16	0.13	0.12	0.08	0.06	0.07	0.07	2.33	9.12
EI	MF/HF	.161	0.23	0.20	0.14	0.14	0.12	0.07	0.06	0.05	2.40	7.78
EE	HF	.171	0.19	0.21	0.16	0.13	0.11	0.09	0.06	0.05	2.52	7.52
AW	MB	.190	0.34	0.08	0.09	0.10	0.10	0.08	0.11	0.10	2.72	9.62
AW	MB	.196	0.34	0.09	0.09	0.09	0.11	0.09	0.10	0.08	2.61	9.40

TABLE 15. PARAMETERS DERIVED FROM REFERENCE PATTERNS FOR MALES WITH IPMOD2-TYPE PREPROCESSING FROM DATA BASE 6.

VOWL	TYP	EER	% 0	% 1	% 2	% 3	% 4	% 5	% 6	% 7	I*P(I)	**
OX	MB	.076	0.25	0.10	0.11	0.15	0.16	0.09	0.07	0.06	2.71	7.70
IR	HC	.084	0.31	0.14	0.11	0.10	0.09	0.09	0.09	0.08	2.54	9.29
UX	HB	.088	0.28	0.16	0.12	0.10	0.10	0.09	0.08	0.05	2.46	8.63
UU	HB	.089	0.29	0.14	0.11	0.10	0.10	0.10	0.09	0.06	2.57	8.78
AU	LB/HB	.091	0.31	0.12	0.10	0.09	0.09	0.10	0.11	0.08	2.69	9.37
EH	MF	.098	0.29	0.19	0.12	0.10	0.10	0.06	0.06	0.07	2.33	8.98
OI	MB/HF	.103	0.26	0.15	0.12	0.11	0.10	0.10	0.10	0.05	2.59	8.42
AE	LF	.104	0.23	0.19	0.14	0.11	0.11	0.09	0.07	0.05	2.50	8.03
AA	LB	.114	0.32	0.08	0.08	0.10	0.13	0.10	0.10	0.09	2.79	9.13
AI	LB/HF	.121	0.37	0.10	0.06	0.05	0.07	0.10	0.14	0.10	2.74	10.43
UH	MC	.128	0.28	0.17	0.12	0.11	0.11	0.09	0.07	0.06	2.46	8.63
IX	HF	.135	0.28	0.15	0.13	0.13	0.10	0.06	0.07	0.08	2.49	8.70
AW	MB	.160	0.30	0.08	0.10	0.12	0.14	0.08	0.10	0.09	2.80	8.73
EE	HF	.173	0.17	0.20	0.14	0.12	0.12	0.13	0.06	0.05	2.73	7.26
EI	MF/HF	.188	0.22	0.18	0.13	0.13	0.12	0.10	0.06	0.06	2.58	7.81
AW	MB	.188	0.33	0.09	0.08	0.09	0.15	0.11	0.07	0.09	2.66	9.11

TABLE 16. PARAMETERS DERIVED FROM REFERENCE PATTERNS FOR FEMALES WITH IPMOD2-TYPE PREPROCESSING FROM DATA BASE 6.

VOWL	TYP	EER	% 0	% 1	% 2	% 3	% 4	% 5	% 6	% 7	I*P(I)	**
UU	HB	.078	0.25	0.17	0.12	0.11	0.10	0.10	0.09	0.05	2.55	8.33
UH	MC	.099	0.22	0.20	0.14	0.11	0.10	0.10	0.10	0.03	2.52	7.91
IR	HC	.102	0.30	0.14	0.10	0.11	0.08	0.08	0.13	0.06	2.58	9.06
OI	MB/HF	.108	0.23	0.13	0.14	0.14	0.13	0.10	0.08	0.04	2.62	7.51
IX	HF	.123	0.21	0.17	0.12	0.12	0.13	0.13	0.09	0.03	2.71	7.43
OX	MB	.132	0.27	0.11	0.10	0.13	0.16	0.09	0.09	0.07	2.73	8.18
AU	LB/HB	.134	0.29	0.15	0.10	0.10	0.10	0.09	0.12	0.05	2.56	8.80
AA	LB	.149	0.37	0.10	0.05	0.08	0.07	0.13	0.12	0.09	2.70	10.19
UX	HB	.159	0.25	0.19	0.12	0.11	0.09	0.09	0.10	0.05	2.49	8.41
AI	LB/HF	.162	0.31	0.15	0.08	0.06	0.10	0.08	0.18	0.04	2.66	9.43
EH	MF	.165	0.20	0.14	0.14	0.13	0.15	0.13	0.09	0.03	2.80	6.97
EE	HF	.179	0.12	0.15	0.14	0.14	0.21	0.18	0.05	0.02	3.02	5.55
AW	MB	.192	0.39	0.07	0.05	0.07	0.04	0.13	0.14	0.10	2.74	10.68
AW	MB	.203	0.30	0.15	0.08	0.07	0.10	0.11	0.15	0.04	2.62	9.08
EI	MF/HF	.206	0.13	0.13	0.13	0.18	0.18	0.13	0.10	0.02	3.03	5.87
AE	LF	.212	0.19	0.19	0.14	0.10	0.12	0.12	0.13	0.03	2.76	7.51

** [P(I)*(1+INT(3.5-I))**2]

TABLE 17. CENTER FREQUENCIES FOR THE BISS FILTERS AND FOR THE CENTRAL VOWEL OF THE PROMPTING WORDS
(FROM PETERSON & BARNEY AND HOLBROOK & FAIRBANKS)

VOWL	BISS EER	IPMOD2 EER	CENTER FREQUENCIES OF FORMANTS (OR FILTERS)
FILT	--	--	410 575 741 906 1071 1237 1402 1567 1733 1898 2063 2229 2394 2559 2725 2890
UU	.105	.089	300 870 2240
UX	.110	.088	440 1020 2240
OI	.116	.103	552 835 2525
			600 1908 2290
AU	.118	.091	770 1400 2695
			610 888 2240
UH	.121	.128	640 1190 2390
OX	.123	.076	495 960 2495
AE	.133	.104	660 1720 2410
EH	.136	.098	530 1840 2480
IR	.137	.084	490 1350 1690
AI	.140	.121	750 1280 2730
			572 1942 2545
AA	.141	.114	730 1090 2440
IX	.161	.135	390 1990 2550
EI	.161	.188	550 2032 2650
			400 2228 2710
EE	.171	.173	270 2290 3010
AW	.190	.160	570 840 2410
AW	.196	.188	

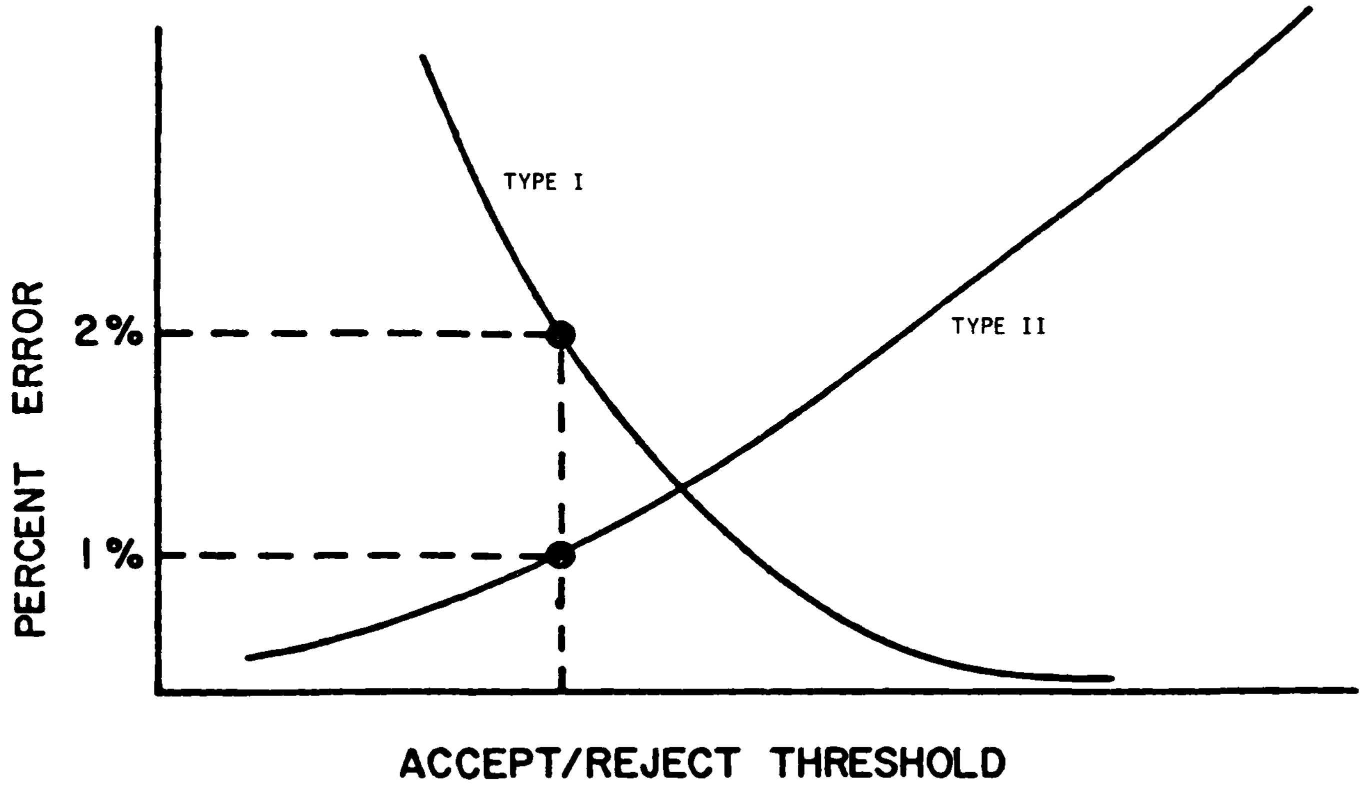


Figure 1. Example of Error Rates versus Threshold.

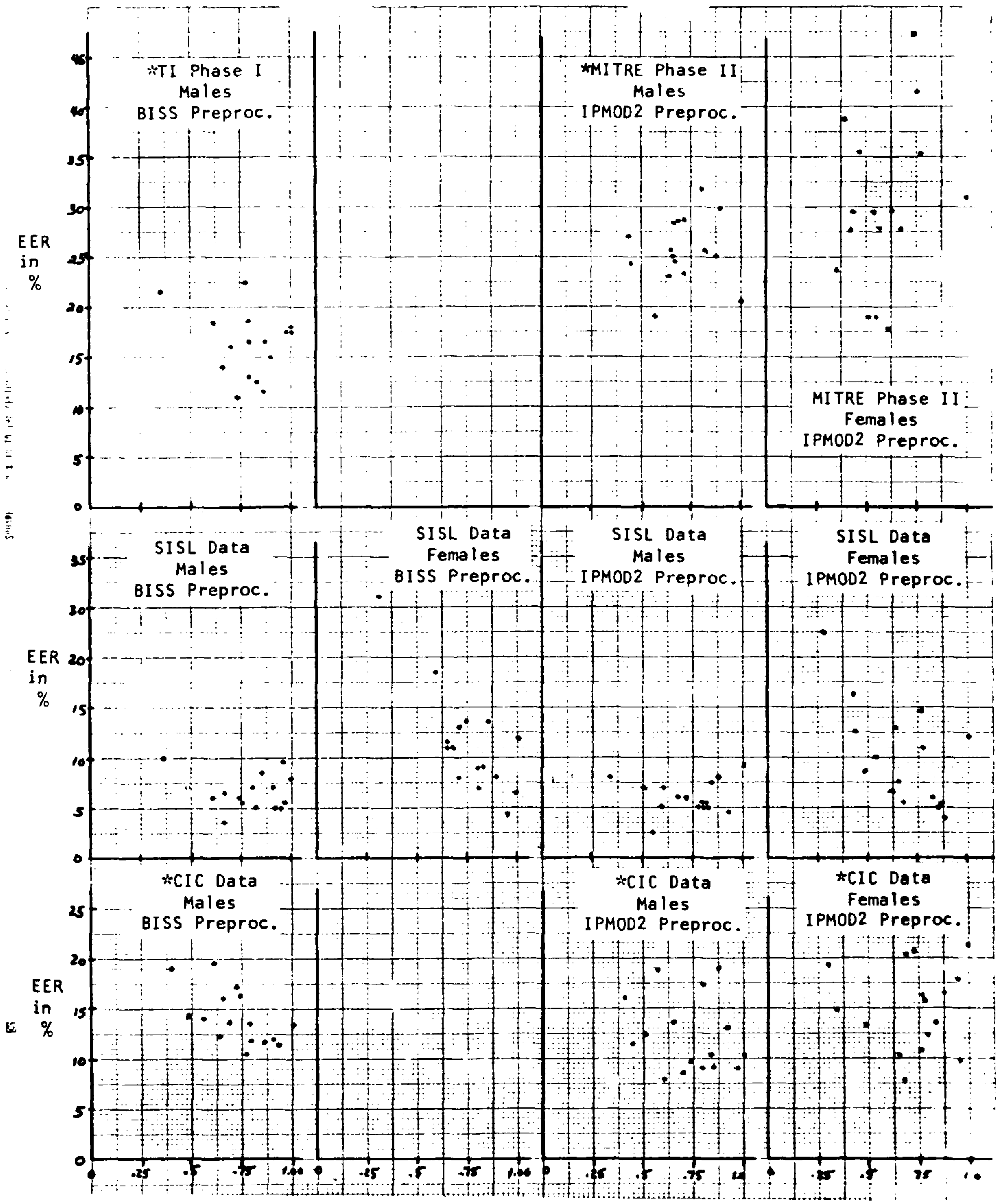


Figure 2 Normalized Average Squared Distances between True Speaker Inputs and References Conditional on Word (16 dots)

AD-A118 444 TEXAS INSTRUMENTS INC DALLAS F/G 17/2
VOICE VERIFICATION UPGRADE.(U)
JUN 82 R L DAVIS, J T SINNAMON, D L COX F30602-79-C-0190
UNCLASSIFIED TI-08-82-07 RADC-TR-82-139 NL

3 of 3

END
DATE
FILMED
09-82
DTIC

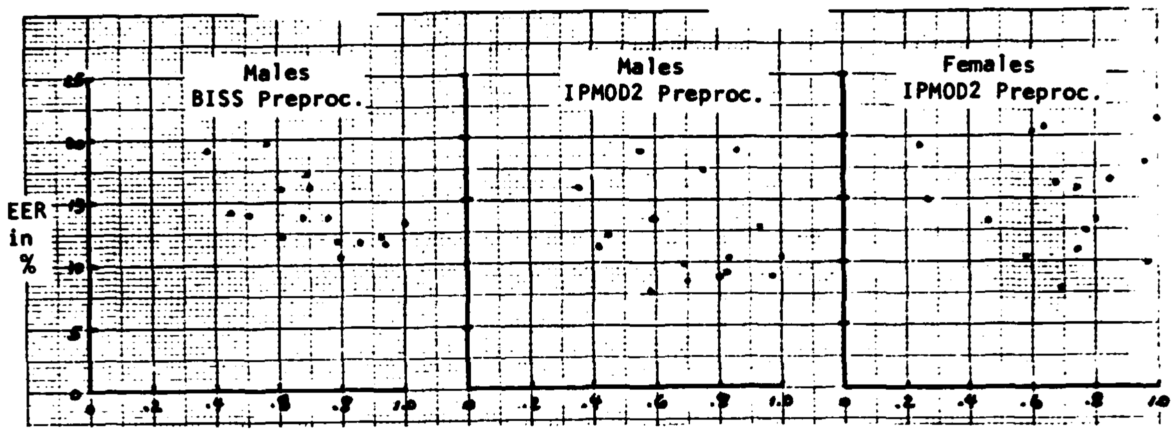


Figure 3 EER vs Medians of True Speaker Distribution for CIC Data (Set 5) Conditional on Word (16 dots)

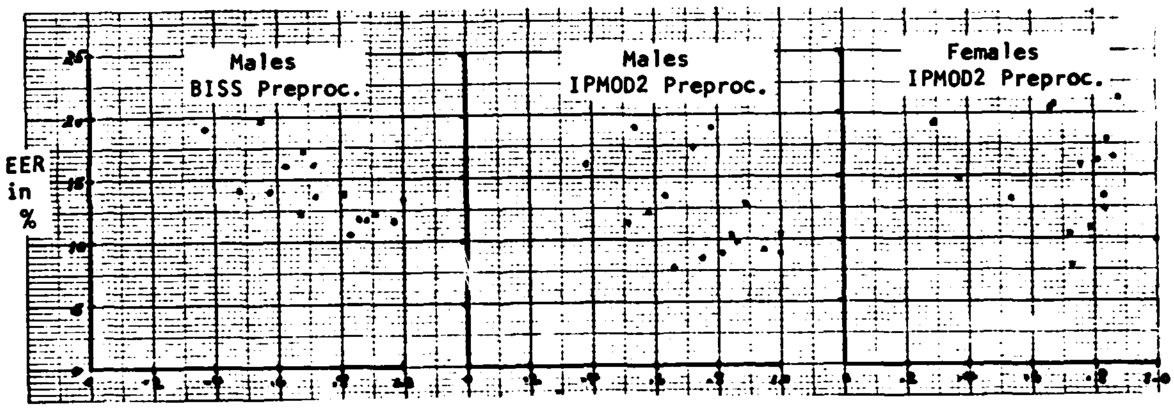


Figure 4 EER vs Equal Error Thresholds for CIC Data (Set 5) Conditional on Word (16 Dots)

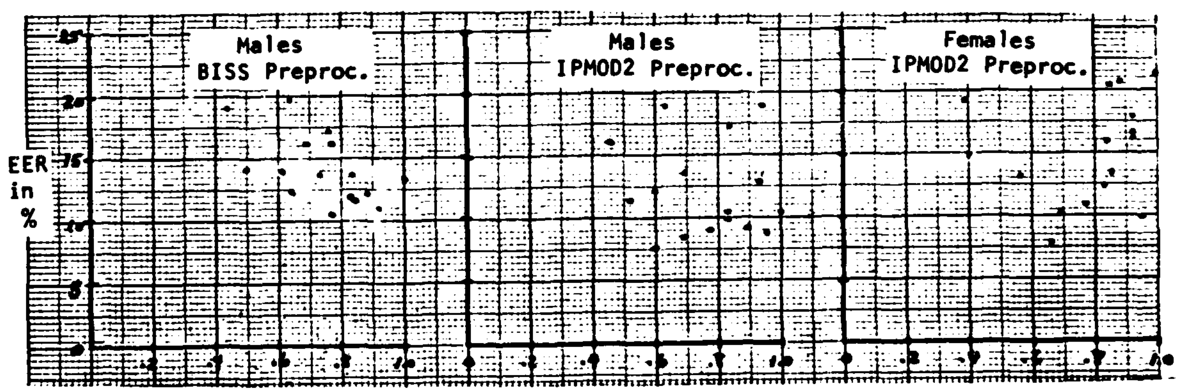


Figure 5 EER vs 10th Percentile of True Speaker Squared Distances for CIC Data Set (5) Conditional on Word (16 dots)

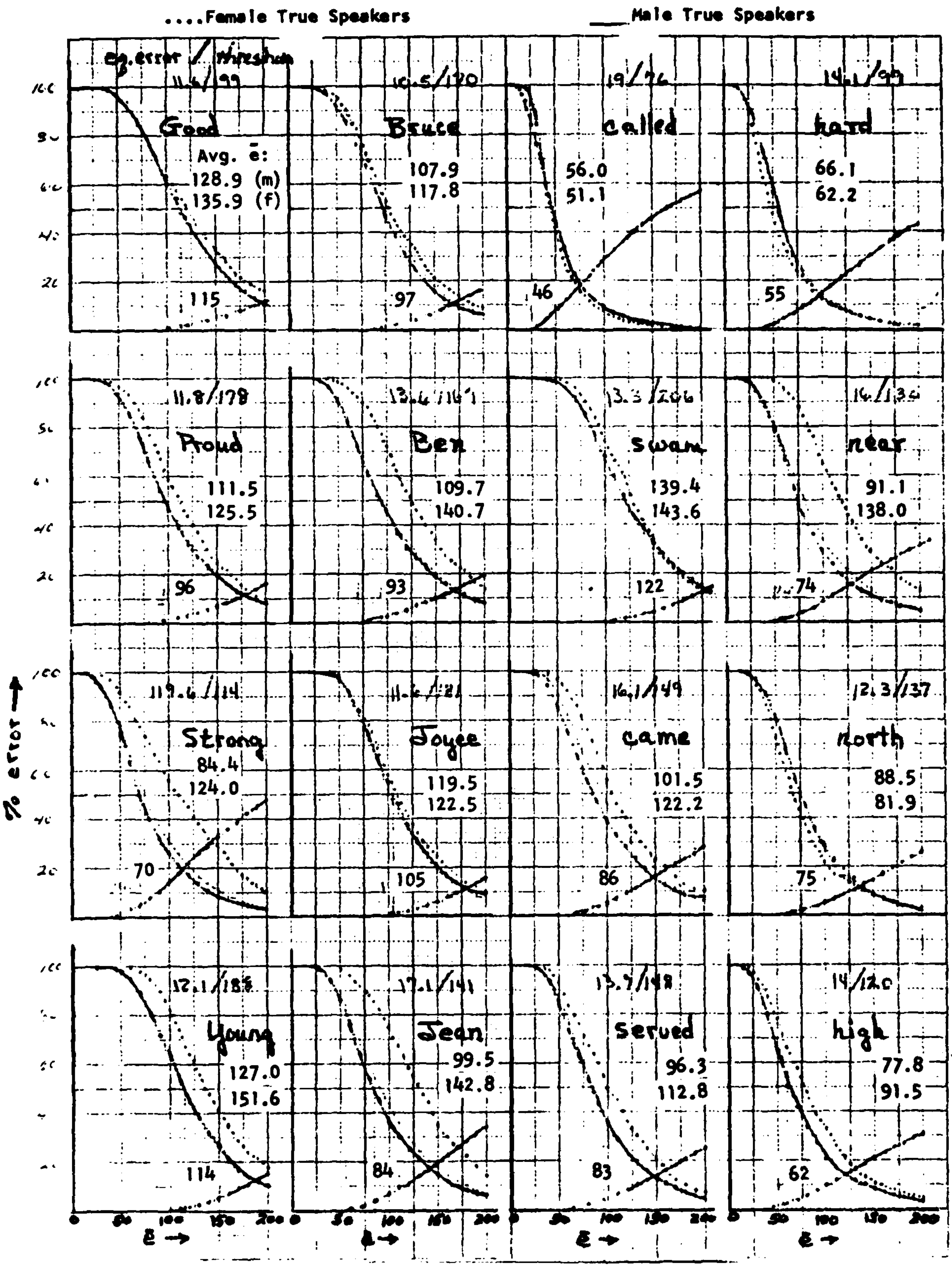


Figure 6 CIC Error Distributions by Word for Males, BISS Preprocessing

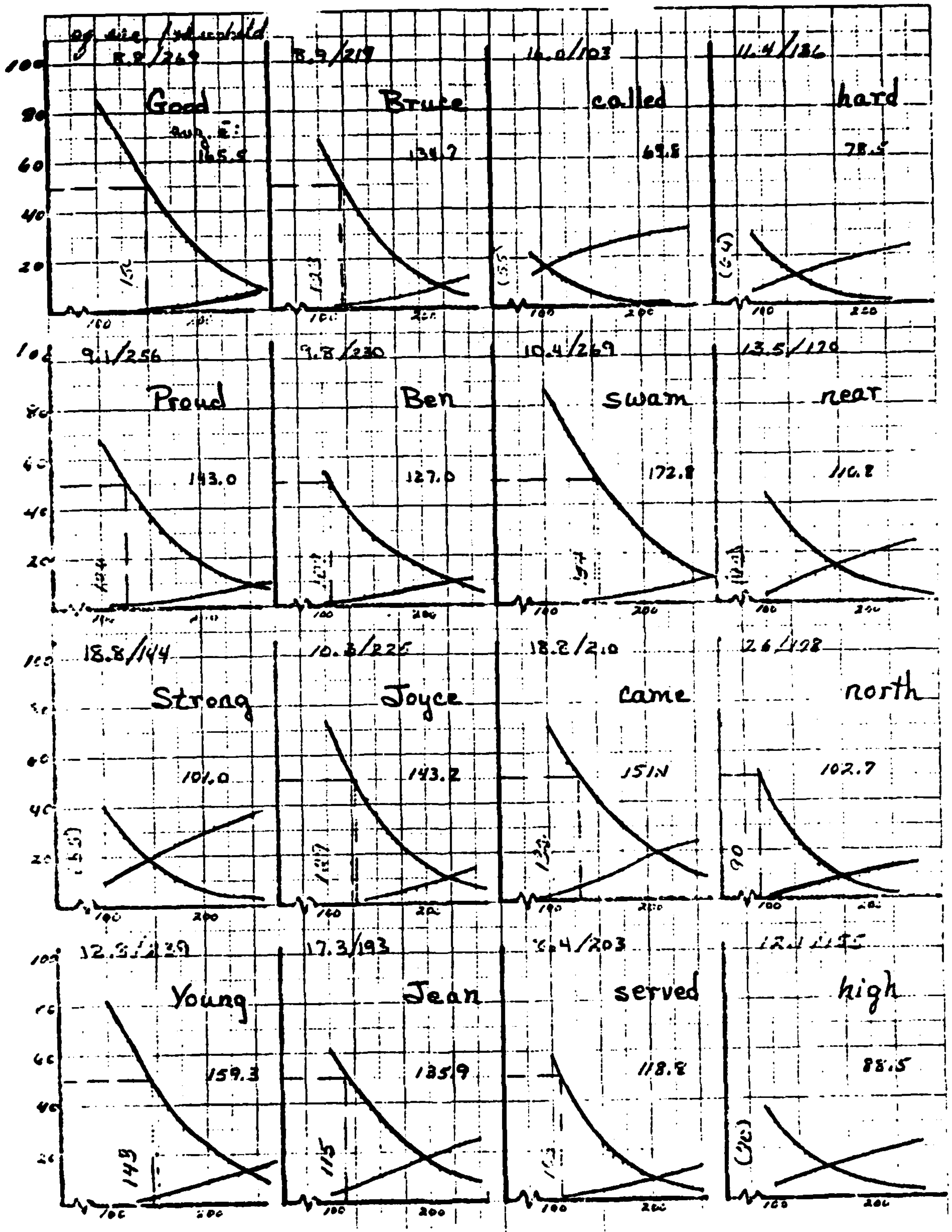


Figure 7 CIC Error Distributions, Male Speakers, IPMOD2 Preprocessing

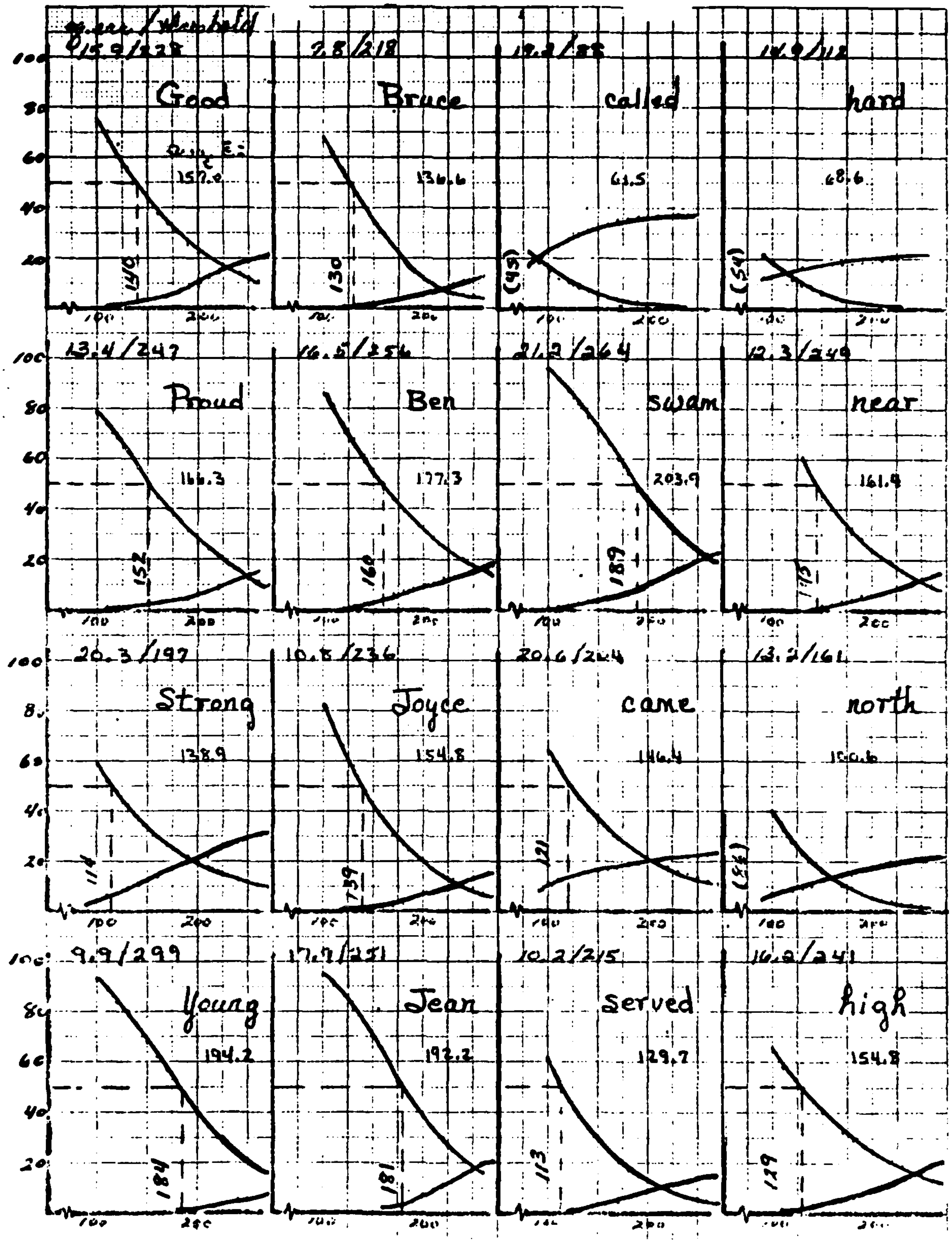


Figure 8 CIC Error Distributions, Female Speakers, IPMOD2 Preprocessing

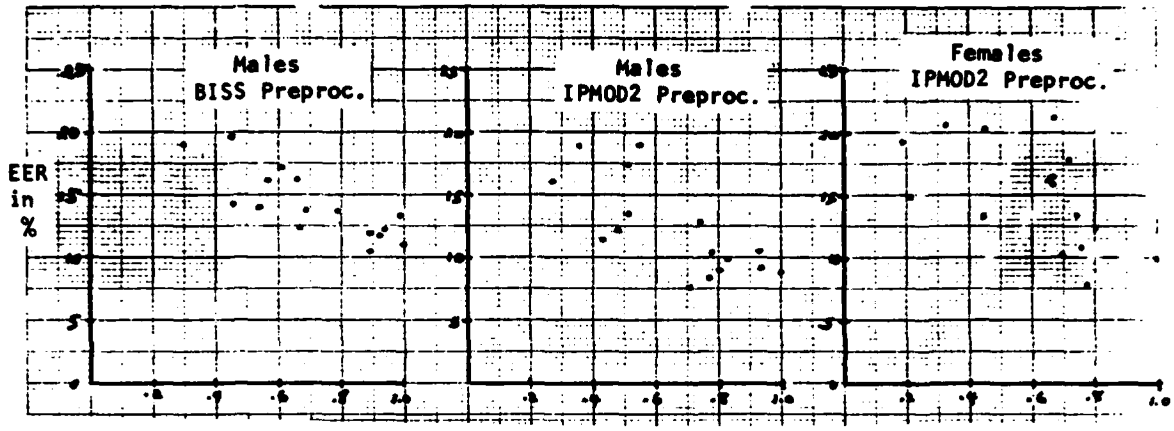


Figure 9 EER vs 10th Percentile of Impostor Square Distances for CIC Data (Set 5) Conditional on Word (16 dots)

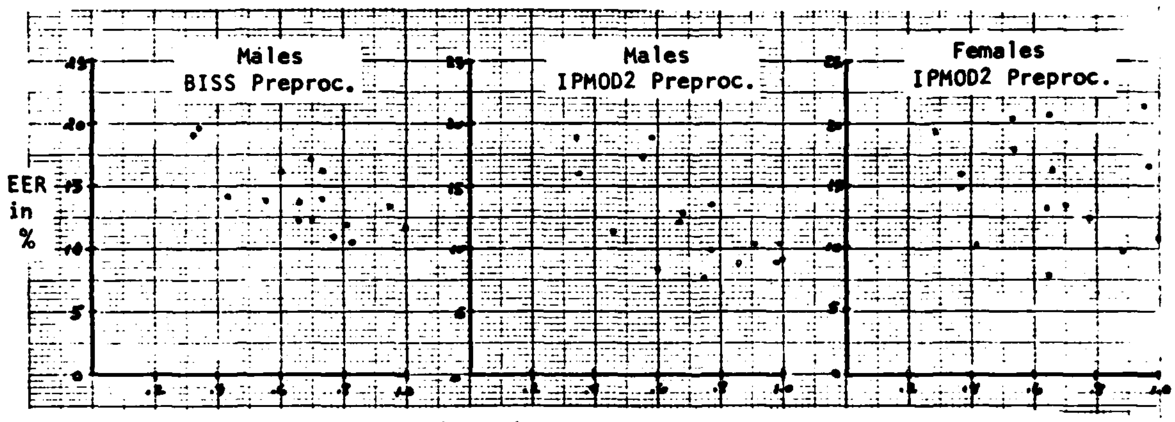


Figure 10 EER for CIC Data (Set 5) vs 30th Percentile of Distances between Reference Files and Average Reference File for Each Word (16 dots) Across all Speakers for CIC Data (set 6)

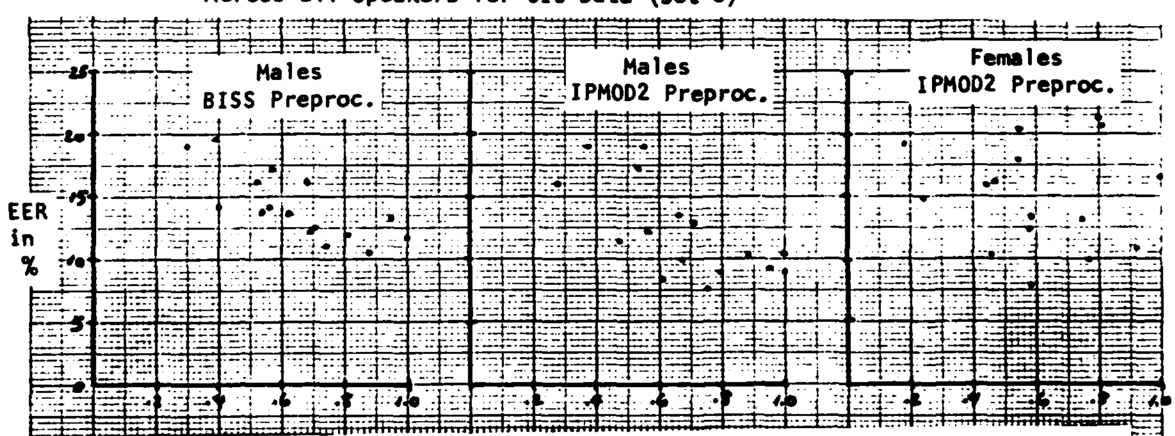


Figure 11 EER for CIC Data (Set 5) vs 10th Percentile of Distances Between all Reference Pattern Pairs for Each Word (16 dots) for CIC Data (Set 6)

Appendix III

ENROLLMENT AND VERIFICATION ALGORITHMS

Appendix III

ENROLLMENT AND VERIFICATION ALGORITHMS

For the algorithms described in this section one of 32 unique phrases is prompted through the prompting unit to the entrant who in turn repeats the phrase into the microphone. The data from the microphone is filtered and compressed and then processed to find either peaks in speech energy or valleys in scanning errors (where the scanning error is the difference between the expected and the received speech patterns). From the peaks or valleys an optimum sequence of 4 peaks or valleys is found and used to create or update reference file data (i.e., expected data) and, in the case of verification, determine if the entrant is verified or not verified. This prompting of phrases and processing of the results is repeated until, in the case of enrollment, a suitable reference file is created, or, in the case of verification, a pass/fail decision is made for the entrant. The following are some of the data which shall be required for these algorithms. Subsequent sections describe the enrollment algorithm, 10.2; the verification algorithm 10.3; the training mode, 10.4; and the validation mode, 10.5.

10.1 Data requirements

10.1.1 Authorization file. The authorization file contains the individual's personal data (name, rank, organization, identification (ID) number and personal code) as well as information pertinent to the operation of the ECS (such as sites to which access is allowed, times/days of authorized entry, escort privilege). This file is created at the time of enrollment and is used during verification to determine if the verification process should take place based on the authorization afforded the individual associated with the ID number imbedded in the badge.

10.1.2 Speech reference file. The speech reference file contains data pertinent to the speech attributes of the individual and consists of expected energy patterns, time intervals and scanning errors for each of the possible words used to compose a phrase. A detailed description of how these data are obtained and used by the system is provided below.

10.1.3 Prompt data. The requirements for the generation of phrases to be prompted through the prompting unit are described herein.

10.1.3.1 Prompt data requirements.

a. Phrases to be prompted to the individual for the purpose of creating a speech reference file and for the purpose of verifying against said reference file data shall consist of phrases listed in Table V.

TABLE V

PROMPTED PHRASES

Phrase Number	Group Number	Phrase	Word Numbers (m)
1	1	north lawn great camp	1, 5, 9, 13
9	1	south limb wide point	2, 6, 10, 14
17	1	east run good plot	3, 7, 11, 15
25	1	first room west cube	4, 8, 12, 16
2	2	north lawn great cube	1, 5, 9, 16
10	2	south limb wide plot	2, 6, 10, 15
18	2	east run good point	3, 7, 11, 14
26	2	first room west camp	4, 8, 12, 13
3	3	north lawn good plot	1, 5, 11, 15
11	3	south limb west cube	2, 6, 12, 16
19	3	east run great camp	3, 7, 9, 13
27	3	first room wide point	4, 8, 10, 14
4	4	north lawn good point	1, 5, 11, 14
12	4	south limb west camp	2, 6, 12, 13
20	4	east run great cube	3, 7, 9, 16
28	4	first room wide plot	4, 8, 10, 15
5	5	north limb wide point	1, 6, 10, 14
13	5	south lawn great camp	2, 5, 9, 13
21	5	east room west cube	3, 8, 12, 16
29	5	first run good plot	4, 7, 11, 15
6	6	north limb wide plot	1, 6, 10, 15
14	6	south lawn great cube	2, 5, 9, 16
22	6	east room west camp	3, 8, 12, 13
30	6	first run good point	4, 7, 11, 14
7	7	north limb west cube	1, 6, 12, 16
15	7	south lawn good plot	2, 5, 11, 15
23	7	east room wide point	3, 8, 10, 14
31	7	first run great camp	4, 7, 9, 13
8	8	north limb west camp	1, 6, 12, 13
16	8	south lawn good point	2, 5, 11, 14
24	8	east room wide plot	3, 8, 10, 15
32	8	first run great cube	4, 7, 9, 16

b. The instructive phrases "louder please" and "thank you" shall be available for prompting to the individual under conditions to be specified below.

c. Other phrases deemed necessary to make the system more pleasant may be added. Some examples to be considered are: "verified", "not verified", "call for assistance", "good morning", etc.

10.1.3.2 Phrase Construction. The phrases to be prompted to the individual for the creation of speech reference file data and for the verification against said reference file data shall be limited to the 32 phrases of Table V. These 32 phrases consist of four words each (n = 1, 2, 3, 4) selected from a total of 16 possible words (m = 1, 2,...16) where the order of occurrence of the four words within any phrase shall be fixed. The use of the index n to represent the position of the word within the phrase as opposed to m which represents the number of the word spoken will be maintained throughout this appendix. As an example, the word "lawn" is word number m=5 but is always spoken as the second word in the phrase, hence n=2. These phrases may be stored either as 32 four word phrases or as 16 words and the phrases constructed at the time of prompting.

10.1.3.3 Order of phrase prompting. The four-word phrases of Table V shall be presented to the individual in a random order where the manner in which these phrases are reordered shall be as follows:

a. Each of the 32 phrases shall be assigned to the group number designated in Table V. Hence each group will consist of a unique set of four phrases.

b. Within each group, the four phrases shall be randomly reordered such that each phrase appears exactly once.

c. The groups shall be randomly reordered such that each group appears exactly once.

d. The multiplicative congruential method of pseudo-random number generation shall be used to reorder the eight groups and the four phrases within each group (reference: D. Knuth, "the art of computer programming" Vol II).

10.1.4 Preprocessing data. The algorithms described herein require that some form of preprocessing be done to the signal at the output of the microphone. As a minimum this shall include all of the filtering described in 3.7.1.1.1.2.2a(1) and may encompass some or all of the processing described in 10.2.2.2.1. Therefore, the preprocessing function shall provide filtered data (which may

additionally be regressed, normalized and/or quantized) from fourteen filters every centisecond (10 milliseconds) in time after the completion of a phrase prompt. The preprocessing function shall also make available a filter overload indication which shall be set whenever filter saturation occurs and which shall remain set until reset as indicated within the algorithms. Additionally, the gain of the filters shall be adjustable as indicated in the algorithms.

10.1.5 Precision requirements. Unless otherwise specified, the precision requirements indicated herein shall be maintained throughout all operations in the algorithms described. The algorithms are specified assuming the precision afforded by a 16-bit computer using integer arithmetic. This degree of precision shall be maintained throughout.

Unless otherwise noted all operations may be truncated to the nearest integer. In those cases where rounding shall be necessary a notation "rounded" will appear in the margin. In some instances significance can be lost if intermediate results are not maintained to the double precision level (i.e. 32-bit integer). These cases will be denoted by "d.p." in the margin and shall be evaluated using enough bits such that neither overflow nor underflow occurs.

10.2 Enrollment algorithm. The enrollment algorithm shall consist of four phases: (1) initialization, (2) generation of initial reference data, (3) refinement of reference data, and (4) termination. Following initialization, phrases are selected and prompted such that speech reference file data is generated for each of the 16 possible words which form the 32 possible phrases. This initial reference file data is then refined by the prompting and processing of phrases a sufficient number of times such that each word is updated a minimum of four times. If this subsequent processing reveals that the initial reference file data was not representative of the individual normal speech, the words involved will be reinitialized and the refinement process repeated. When all words have been sufficiently refined, the resultant reference file data will be saved for the verification attempts which the entrant will be subsequently undertaking.

During the generation and refinement of the reference file data, the enrollment operator or the enrollee shall be allowed to interrupt the procedures in the manner described in 10.2.5. With this exception, the enrollment shall proceed in the sequence described below. Figure 6 is a flowchart of the enrollment algorithm where the paragraph numbers in the figure correspond to those in the text. The notation glossary of Table VI will be adhered to in the following description.

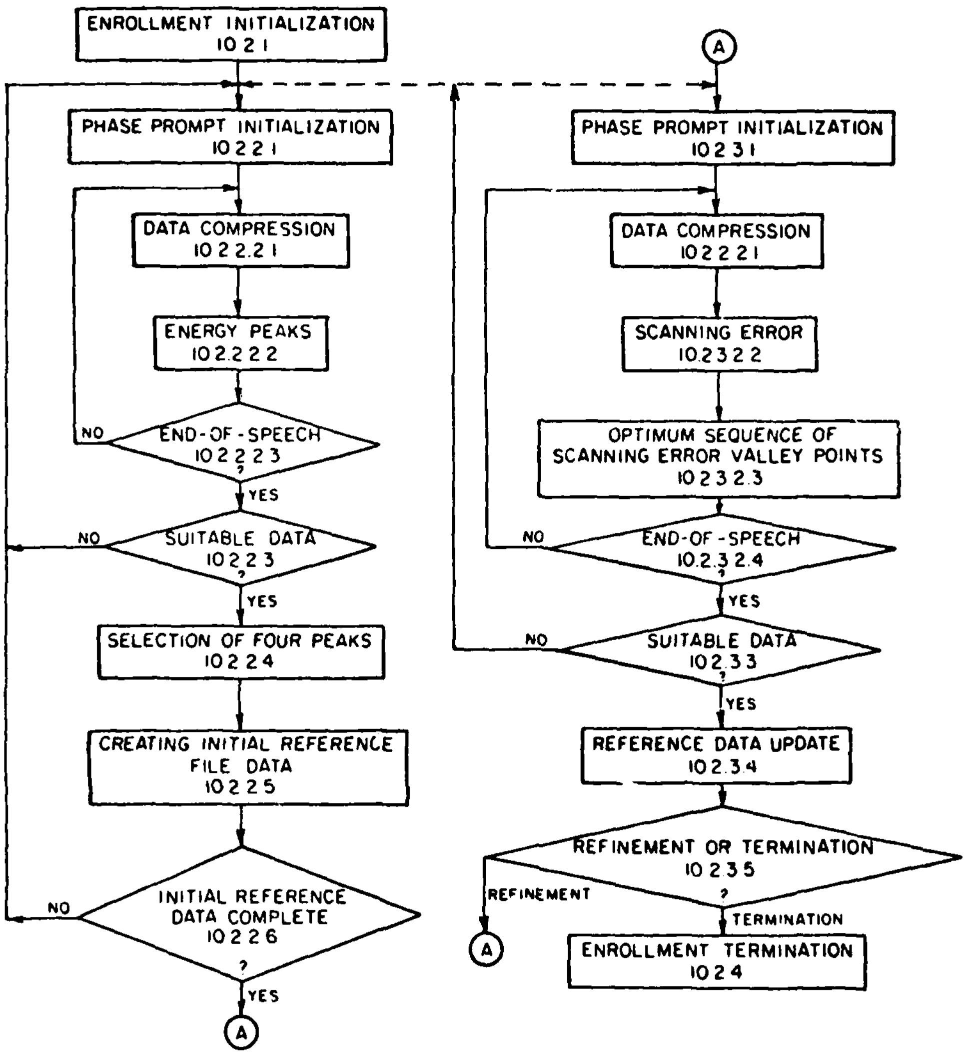


Figure 6. ENROLLMENT ALGORITHM

TABLE VI

ALGORITHM NOTATION GLOSSARY

Symbol	Usage	Major Reference
α	Reference file updating parameter	10.3.3.1.3
BESTSQ	Optimum sequence parameter	10.2.2.4, 10.2.3.2.3.2.3
C_1, C_2	Regression coefficients	10.2.2.2.1.1
e(L)	Maximum filter energy at time L	10.2.2.2.1.1
$\hat{e}$	A decision score parameter	10.3.3.1.2
EHAT	Summed expected scanning error	10.3.3.1.1
$ep(j_n)$	Energy peak j for the nth word	10.2.2.2.2
EPEAK	Peak energy	10.2.2.2.3.1
es(L)	Smoothed energy at time L	10.2.2.2.2
ESE_n, $ESE_{m(n)}$	Reference data-expected scanning error for word m	10.2.3.1, 10.2.4, 10.3.3.1.3
eseq	Sequence error	10.2.2.4, 10.2.3.2.3.2.3
$ESMIN_n(j_n)$	Scanning error of the jth valley point for the nth word	10.2.3.2.3.1
ETH_m, $ETH_{m(n)}$	Initial expected relative energy of word m	10.2.2.4
ETOT	Total energy in sequence of peaks	10.2.2.4
EUSE	Summed scanning error	10.3.3.1.1
EW	Point pair error for a valley point couplet	10.2.3.2.3.2.1.2
f(i)	Output from filter i	10.2.2.2.1
g(i)	Regressed output from filter i	10.2.2.2.1.1

$\gamma_1(i), \gamma_2(i)$	Regression vectors	10.2.2.2.1.1
h(i)	Normalized output from filter i	10.2.2.2.1.2
i	Filter index ($1 \leq i \leq 14$)	10.2.2.2.1
I	Valley sequence index	10.2.3.2.3.2
IREG	Registered phrase counter	10.3.3.1.1
j	General purpose index	-----
j_n	Index to the jth peak or valley for the nth word	10.2.2.4, 10.2.3.2.3.2
k	Time sample index	-----
k_I	Index to the kth couplet of valley points for the Ith and (I+1)th words	10.2.3.2.3.2
K	Filter quantization parameter ($1 \leq K \leq 7$)	10.2.2.2.1.3
L	Time in centiseconds since end of prompt	-----
LTIME	Parameter used to locate energy peaks	10.2.2.2.2
$LTIME_n$	Parameter used to locate valley points for the nth word	10.2.3.2.3.1
m, m(n)	Word number ($1 \leq m \leq 16$) where m(n) emphasizes the fact that word m is spoken as the nth word in the phrase	------
MAX	Parameter used to locate energy peaks	10.2.2.2.2
MAXCUT	Expected length of speech	10.3.2.3
MIN_n	Parameter used to locate valley points for the nth word	10.2.3.2.3.1
MODE	Parameter used to locate energy peaks	10.2.2.2.2

$MODE_n$	Parameter used to locate valley points for the nth word	10.2.3.2.3.1
MXD	Parameter used to compute point pair errors	10.2.3.2.3.2.1.2
n	Position of a word within a phrase ($1 \leq n \leq 4$)	-----
$\eta(i,L)$	Quantized output for filter i at time L	10.2.2.2.1.3
NOTREG	Mis-registered phrase counter	10.2.3.3, 10.3.3.3
NPK	Energy peak counter	10.2.2.2.2, 10.2.2.4
NTC_m, $NTC_{m(n)}$	Number of times reference data for word m has been calculated	10.2.2.5, 10.2.3.1, 10.2.3.4
$OPTSE_n$	Scanning error of the nth valley point or energy of the nth peak in the optimum sequence	10.2.3.2.3.2.3
$OPTT_n$	Time of the nth valley point or the nth peak in the optimum sequence	10.2.3.2.3.2.3
PDEC	Decision score	10.3.3.1.2
PPE	Scanning errors of couplets for the valley points from words I and I+1	10.2.3.2.3.2.1.2
PPEW	Point pair errors of couplets for the valley points from words I and I+1	10.2.3.2.3.2.1.2
PPT	Times of couplet for the valley points from words I and I+1	10.2.3.2.3.2.1.2
$\phi_1(i)$, $\phi_2(i)$	Regression vectors	10.2.2.2.1.1
$\psi(i,k)$	Quantization thresholds	10.2.2.2.1.3
$r_m(i,k)$	Reference data - expected quantized filter i output for time sample k, word m	10.2.3.1, 10.2.4 10.3.3.1.3

$SE_n(L)$	Scanning error for the nth word at time L	10.2.3.2.2
sESE	Summed expected scanning error for a phrase	10.3.2.1, 10.3.3.1.1
$sESE_m$	Summed expected scanning error for word m	10.2.2.5, 10.2.3.4
$SERR_n$	Parameter used to compute sequence error	10.2.2.4
$sr_m(i,k)$, $sr_{m(n)}(i,k)$	Summed pattern data for filter i, time sample k and word m	10.2.2.5, 10.2.3.4
ST	Parameter used to compute sequence error	10.2.2.4
$s\Delta Tref_m$, $s\Delta Tref_{m(n)}$	Summed expected time interval between word m prompted as the nth word and the (n+1)th word	10.2.2.5, 10.2.3.4
SPEECH	Start of speech indicator	10.2.2.2.3.1
t	General purpose parameter representing time	-----
ΔT	Time interval between valley points	10.2.3.2.3.2.1.1
Δte_m, $\Delta te_{m(n)}$	Initial expected time interval between word m prompted as the nth word and the (n+1)th word	10.2.2.4
$tp\ (j_n)$	Time of the jth peak for the nth word	10.2.2.2.2
$\Delta Tref_m$, $\Delta Tref_{m(n)}$	Reference data - expected time interval between word m prompted as the nth word and the (n+1)th word	10.2.3.1, 10.2.4, 10.3.3.1.3
$TSMIN_n(j)$	Time of the jth valley point for the nth word	10.2.3.2.3.1
μ	Filter normalization parameter	10.2.2.2.1.2
$udt_j\ (n)$	Updating time interval between	10.3.3.1.1

	the nth and (n+1)th words of the jth registered phrase	
$uptn_j(i,k,n)$	Updating pattern data for filter i, time sample k for the nth word of the jth registered phrase	10.3.3.1.1
$use_j(n)$	Updating scanning error data for the nth word of the jth registered phrase	10.3.3.1.1
$uwd_j(n)$	Number of the word prompted as the nth word of the jth registered phrase	10.3.3.1.1
x	General purpose variable	-----

10.2.1 Enrollment initialization. Initialization shall consist of the following:

a. The enrollment function shall obtain a unique identification (ID) number for the enrollee which shall associate the authorization file with the speech reference file (generated here) through the coded badge. This ID number shall also be available to the enroller.

b. The filter gain G, 3.7.1.1.1.1.3, which has possible values of 1, 2, 4,... up to a maximum filter gain shall be initialized to two and the filter overload indicator shall be cleared.

c. The number of times reference data has been calculated for each of the 16 possible words shall be set to zero, i.e., $NTC_m = 0$ for m = 1, 2,...16.

d. A list of 32 phrases (eight groups) in random order shall be created as described in 10.1.3.3. When the actual enrollment process begins, the first phrase in this list shall be the phrase which is prompted to the enrollee.

Upon completion of initialization the procedures of 10.2.2 shall be followed.

10.2.2 Generation of initial reference file data. The generation of initial reference file data consists of locating peaks in the speech energy and selecting from those peaks four peaks corresponding to the four spoken words such that the time separation between peaks falls within expected intervals. The data selected to initialize the reference file is: (1) the data compressed filter outputs centered at the time of the peaks and (2) the time intervals between the peaks. This process is repeated until data has been generated in this manner for each of the 16 possible words used to create all phrases.

10.2.2.1 Phrase prompt initialization. Prior to the processing of speech data for a given prompted phrase, the following initialization shall be performed.

a. The parameters MAX and LTIME, used to locate peak energies, shall be set to 125 and zero respectively. The mode used to locate peaks shall be set to search-for-valley. The number of peaks found, NPK, shall be set to zero.

b. The parameters EPEAK and SPEECH used to find the end of speech shall be set to zero and not-started respectively.

c. The parameter BESTSQ used to locate an optimum sequence of peaks shall be set to the value of the largest positive integer possible for the computer.

d. The phrase indicated by previous processing shall be prompted through the prompting unit.

Upon completion of this initialization, speech processing shall begin as per 10.2.2.2.

10.2.2.2 Speech processing. For each centisecond following the end of prompting, while the enrollee is repeating the prompted phrase and until end-of-speech is declared, 10.2.2.2.3, the following procedures shall be implemented.

10.2.2.2.1 Data compression. Data compression consists of accepting the filtered data 3.7.1.1.1.2.2.a(1) and regressing, normalizing and quantizing it in the manner described herein. The data to be compressed consist of one data point f (with at least 12 bit quantization) from each of fourteen filters and occuring every centisecond. The output from the data compression shall be the maximum filter energy e(L) and the quantized data $\eta(i,L)$ for i=1, 2,...14 at time L=1,2,... in centiseconds.

10.2.2.2.1.1 Regression. The inputs f(i) for i=1,2,...14 shall be regressed to eliminate slope and curvature of the spectrum in the following manner.

a. If any one of the 14 inputs is less than one, the corresponding f(i) shall be set to one.

b. Regression coefficients shall be computed as follows:

$$C = 7 + \sum_{i=1}^{14} \frac{f(i)*\phi_1(i)}{32768} \qquad \text{(d.p.)}$$

$$C = 7 + \sum_{i=1}^{14} \frac{f(i)*\phi_2(i)}{32768} \qquad \text{(d.p.)}$$

where $\phi_1(i)$ and $\phi_2(i)$ shall be as provided in Table VI.

c. The inputs shall be regressed as follows:

$$g(i) = f(i) + \frac{C*\gamma_1(i)}{32768} \quad \frac{C*\gamma_2(i)}{32768} \qquad \text{(d.p.)}$$

TABLE VII

REGRESSION VECTORS

i	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$	$\phi_4(i)$
1	-32767	32767	6085	-7607
2	-27726	17644	5149	-4096
3	-22685	5041	4212	-1170
4	-17644	-5040	3276	1169
5	-12602	-12602	2340	2925
6	-7561	-17644	1403	4095
7	-2520	-20164	467	4681
8	2521	-20164	-468	4681
9	7562	-17644	-1404	4095
10	12603	-12602	-2341	2925
11	17644	-5040	-3277	1169
12	22685	5041	-4213	-1170
13	27726	17644	-5149	-4096
14	32767	32767	-6086	-7607

for i = 1, 2,...14

where $\gamma_1(i)$ and $\gamma_2(i)$ shall be as provided in Table VI.

d. The maximum filter energy at the current time L shall be taken as the maximum of the fourteen values of g(i) and will be denoted e(L), i.e.,

$$e(L) = \max_{1 \le i \le 14} \{g(i)\}$$

10.2.2.2.1.2 Normalization. The regressed values g(i) shall be normalized as follows:

$$h(i) = (32768 * g(i))/\mu \qquad \text{(d.p., rounded)}$$

where

$$\mu = \sum_{i=1}^{14} g(i)$$

10.2.2.2.1.3 Quantization. The normalized values h(i) shall be quantized as follows:

$$\eta(i,L) = \begin{cases} 0 \text{ for } h(i) \le \Psi(i,1) \\ K \text{ for } \Psi(i,K) < h(i) \le \Psi(i,K+1) & K = 1, 2,\ldots 6 \\ 7 \text{ for } \Psi(i,7) < h(i) \end{cases}$$

where $\Psi(i,K)$ shall be as provided in Table VIII.

10.2.2.2.2 Energy peaks. If 13 centiseconds has not yet elapsed since the end of prompting (i.e., if L < 13), the end-of-speech test as per 10.2.2.2.3 shall be performed. If, however, L $\ge$ 13, the energy peaks of the smoothed energy function shall be located. The smoothed energy es(L) at time L shall be defined as

$$es(L) = \sum_{K=1}^{13} \frac{e(L-13+K)}{2}$$

The locating of peaks in this function shall be done by use of two modes (search-for-valley and search-for-peak) where the mode shall be that which was determined at the immediately previous time sample, time=(L-1), or during initialization if L=13. Based on the

TABLE VIII

QUANTIZATION THRESHOLDS

i	$\Psi(i,1)$	$\Psi(i,2)$	$\Psi(i,3)$	$\Psi(i,4)$	$\Psi(i,5)$	$\Psi(i,6)$	$\Psi(i,7)$
1	2067	2434	2871	3264	3884	4643	5684
2	2074	2625	3185	3713	4247	4712	5549
3	1722	1923	2097	2359	2859	3323	4032
4	1574	1779	1978	2201	2476	3267	4582
5	1536	1754	1920	2110	2404	2764	4302
6	1662	1886	2072	2349	2674	3401	4612
7	1749	1964	2132	2327	2606	3014	3740
8	1774	2020	2227	2438	2758	3329	4367
9	1691	1937	2206	2496	2902	3394	4119
10	1586	1886	2220	2581	3111	3671	4526
11	1569	1854	2137	2469	2869	3354	4399
12	1731	2016	2309	2620	2961	3430	4289
13	1890	2229	2468	2750	3060	3358	3867
14	1858	2163	2425	2763	3059	3416	3901

current mode the values to which the parameters MAX, LTIME, and MODE shall be updated for the next centisecond (at L+1) shall be as presented in Table IX. As an example, if es(L) is less than the current value of MAX and the current mode is search-for-valley, MAX shall be set to es(L), LTIME shall be set to L and the mode for the next time at L+1 shall remain as search-for-valley.

If the current mode is search-for-peak and (MAX*10)/12 is greater than or equal to es(L) then a peak is said to be located. In this case, prior to the updating indicated in the table and then only if the unupdated value of MAX is greater than or equal to 150, the peak shall be saved. If the peak is to be saved, NPK shall be incremented by one and the peak value and location shall be saved in arrays ep and tp respectively as follows:

$$ep(NPK) = MAX/50 \qquad \text{(rounded)}$$

$$tp(NPK) = LTIME$$

Provision shall be made for up to 90 such peaks being found.

Once the energy peak check is complete for the current time, the end-of-speech test as per 10.2.2.2.3 shall be performed.

10.2.2.2.3 End-of-speech. At each centisecond in the speech processing the end-of-speech test described herein shall be performed. The criterion to be met is either one of the following two:

(1) one hundred centiseconds of valid speech exists prior to a speech silence;

(2) speech has continued for 400 centiseconds since the end of prompting.

The testing to determine if one of these two criteria is met shall be as described in 10.2.2.2.3.1 and 10.2.2.2.3.2. If end-of-speech is not declared, speech processing shall continue. If end-of-speech is declared, subsequent processing shall be the determination of whether the data is suitable (i.e. 10.2.2.3 if generation of initial reference file data is being implemented; 10.2.3.3 if refinement of reference file data is being implemented; or 10.3.3 if a verification is being implemented).

10.2.2.2.3.1 Speech duration. The determination of whether end-of-speech should be declared due to the speech duration test rests upon three dependent conditions, Figure 7. First the fact that speech has begun must be ascertained. Having determined that speech has begun, bands of silence of a minimum of 60 centisecond duration are

TABLE IX

ENERGY PEAK TABLE

Current Mode	Condition	MAX	LTIME	MODE
Search for Valley	es(L) < MAX	es(L)	L	*
	MAX ≤ es(L) < (MAX*12)/10	*	*	*
	(MAX*12)/10 ≤ es(L)	es(L)	L	Peak
Search for Peak	es(L) > MAX	es(L)	L	*
	MAX ≥ es(L) > (MAX*10)/12	*	*	*
	(MAX*10)/12 ≥ es(L)	es(L)	L	Valley

*Unchanged

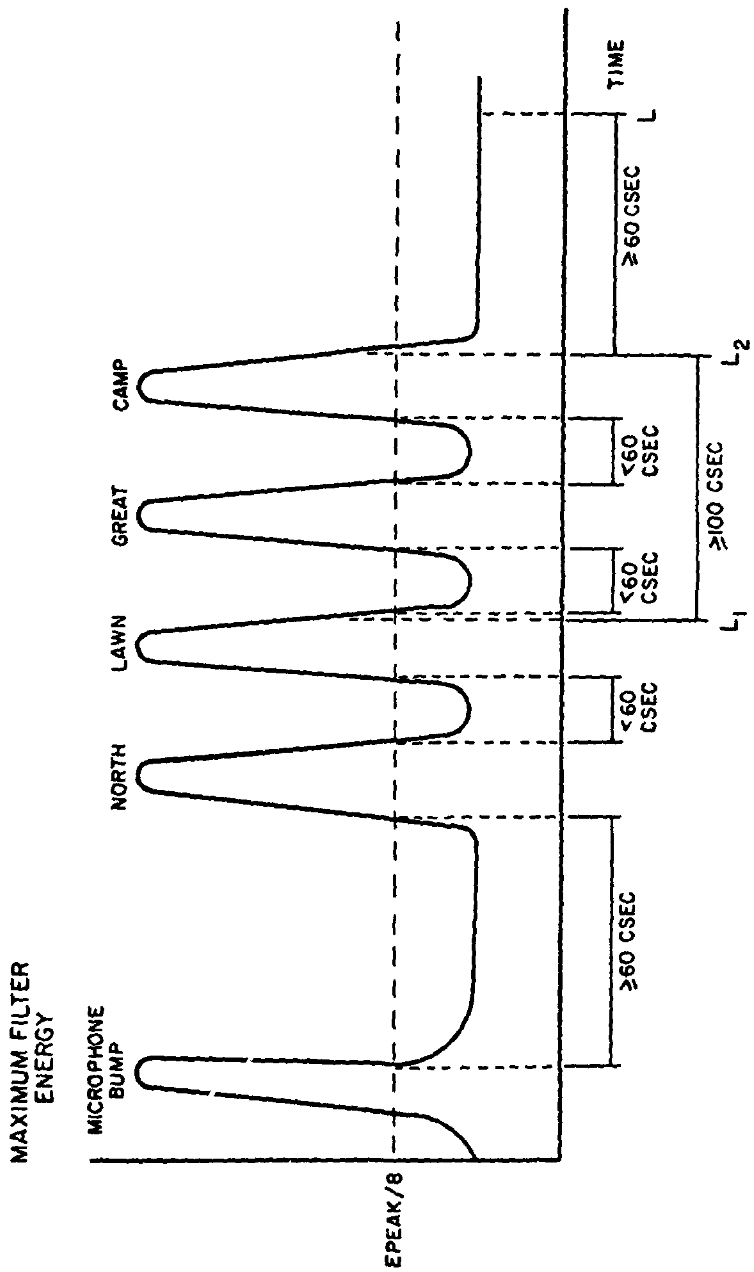


Figure 7. END OF SPEECH AT L > 400

sought. Then, to determine if a given band of silence is indeed at the end-of-speech, the past history of the maximum filter energy function is investigated to determine if valid speech preceded the silence. If any one of these conditions is not met, processing shall proceed to section 10.2.2.2.3.2 with the speech duration test having failed. If all conditions are met, end-of-speech shall be declared and processing shall proceed as indicated previously. The determination of whether these conditions have been made shall be as follows.

10.2.2.2.3.1.1 Speech start. Speech is recognized as having started by the fact that e(t), the maximum filter energy function, has attained a value of 100 or more at some point in time. If speech start has already been recognized as per this test, processing shall proceed to 10.2.2.2.3.1.2. If speech start has not yet been ascertained but at this time L, $e(L) \geq 100$, speech start shall be declared. In doing so, speech shall be set to started and EPEAK shall be set to this value of e(L) as a first step in maintaining EPEAK as the maximum of e(t) for $t \leq L$. Irrespective of whether speech start is declared at this time, processing shall proceed to 10.2.2.2.3.2.

10.2.2.2.3.1.2 Speech silence. Speech silence is said to occur when the maximum energy function e(t) falls below the speech silence threshold EPEAK/8. As EPEAK is to be the maximum of e(t) for $t \leq L$, EPEAK shall be updated to the current value of e(L) if e(L) > EPEAK. If at any time e(L) is seen to fall below the current value of the silence threshold, the time at the previous interval (i.e. L-1) shall be recorded as L_2 and a count shall be started to count the number of consecutive maximum filter energy samples that remain below this threshold. Whenever a span of speech silence is encountered which is 60 centiseconds or more in length (i.e. the count becomes greater than or equal to 60), speech silence of sufficient length is said to exist and processing shall proceed to valid speech, 10.2.2.2.3.1.3. If at the current time, speech silence does not exist or is of insufficient length, processing shall proceed to 10.2.2.2.3.2.

10.2.2.2.3.1.3 Valid speech. To preclude the possibility of accepting as speech some accidental noise (such as is caused by the microphone being bumped), speech is only said to be valid if before speech silence, a span of maximum filter energy can be found which lasts for at least 100 centiseconds and contains no speech silence intervals longer than 59 centiseconds. Hence, once a band of speech silence is located as per 10.2.2.2.3.1.2, the e(t) function for $t < L_2$ (where L_2 is the time before the speech silence) shall be investigated in the following manner. Letting $L_1 = L_2 -1, L_2 -2, \ldots$ values of $e(L_1)$ shall be compared with EPEAK/8. If for some L_1 , $e(L_1) <$ EPEAK/8, a count shall be started to count the number of

consecutive $e(L_1)$ samples of speech silence. If any such count reaches 60, valid speech is not said to exist in which case SPEECH shall be reset to not-started and processing shall proceed to maximum speech duration, 10.2.2.2.3.2. If, however, the current value of $e(L_1) \geq EPEAK/8$, the time length, $L_2 - L_1$, shall be compared to 100. If $L_2 - L_1 \geq 100$, valid speech exists and end-of-speech shall be declared with processing continuing as has been indicated in 10.2.2.2.3. If $L_2 - L_1 < 100$, no determination as to valid speech can be made and previous values of $e(L_1)$ shall be further investigated in the described manner. If no such determination can be made by the time L_1 has been decremented to $L_1=1$, valid speech is said not to exist, in which case SPEECH shall be reset to not-started and processing shall proceed to 10.2.2.2.3.2.

10.2.2.2.3.2 Maximum speech duration. No speech processing shall continue for longer than 400 centiseconds of speech data. If the amount of time since the end of prompting has reached 400 centiseconds (i.e., if L = 400), end-of-speech is summarily declared and speech processing shall cease.

10.2.2.3 Suitable data. Once end-of-speech is declared, the results of the speech processing shall be investigated so as to determine if the data is suitable for use in creating initial reference data. If the data is deemed suitable, the selection of four peaks as per 10.2.2.4 shall be implemented. If the data is deemed not suitable, all data resulting from the processing of the current phrase shall be discarded, this same phrase shall be reprompted as the next phrase, and processing shall be resumed with phrase prompt initialization, 10.2.2.1. This procedure postpones the prompting of the next phrase in the list created at the start of enrollment, 10.2.1.d, until speech processing has yielded data deemed suitable by this test. There shall be no limit to the number of times a given phrase may be reprompted in this manner.

Two conditions may arise, either of which is sufficient to declare the data not suitable. The first condition is a filter overload. If the filters yielded an overload status condition, then in addition to the above, the procedure 10.2.2.3.1 shall be implemented. If no filter overload occurred but fewer than four peaks were found by 10.2.2.2.2, i.e., NPK < 4, then the second condition has been met and the gain adjustment test of 10.2.2.3.2 shall be implemented, where mis-registered shall be defined as having found no more than two peaks, i.e., NPK $\leq$ 2.

10.2.2.3.1 Filter overloads. Whenever a filter overload status is sensed, two messages shall be sent to the preprocessor. The first shall be to cause the filters to be reset. The second shall be to reduce the filter gain, G, by a factor of two to a level no less than one (i.e. if G=1, no further gain reduction is possible).

10.2.2.3.2 <u>Filter gain adjustment</u>. Whenever a new phrase is to be prompted and no filter overload has occurred, the following test shall be performed to determine if the filter gain G shall be adjusted or if the instruction "louder please" shall be prompted to the enrollee. This decision is based on the value EPEAK, the maximum value of e(L), 10.2.2.2.1.1.d, used to determine when end-of-speech was declared.

a. If EPEAK > 1200, G shall be decreased by a factor of two to a level no less than one.

b. If EPEAK $\leq$ 1200 and if the phrase did not mis-register then

(1) If EPEAK/G < 150 and the instruction "louder please" has not yet been prompted during the entry attempt, "louder please" shall be prompted. ("louder please" shall be prompted no more than once per entry attempt.)

(2) If EPEAK < 400 and "louder please" is not to be prompted at this time, G shall be increased by a factor of two to a level no greater than the maximum filter gain.

10.2.2.4 <u>Selection of four peaks</u>. If the speech processing described in yielded data deemed suitable as per 10.2.2.3, then the peak data saved as per 10.2.2.2.2 shall be considered so as to find four peaks representative of the four prompted words in the phrase. Table X presents values of $\Delta te_{m(n)}$ (initial expected time intervals between word m prompted as the nth word in the phrase and whatever word was prompted next) and of $ETH_{m(n)}$ (the expected relative energy of word m, the nth word in the phrase). For the particular words used in the phrase just prompted, the appropriate values of $\Delta te_{m(n)}$ and $ETH_{m(n)}$ shall be used. The peaks of 10.2.2.2.2 were saved in arrays tp(j) and ep(j) where $1 \leq j \leq$ NPK. From these NPK peaks all possible combinations of selecting four peaks j_1, j_2, j_3 and j_4 such that $1 \leq j_1 < j_2 < j_3 < j_4 \leq$ NPK shall be considered. If the combination (or sequence) currently under consideration meets the following time separation criterion:

$$\Delta te_{m(n)}/2 \leq tp(j_{n+1}) - tp(j_n) \leq 2 * \Delta te_{m(n)}$$

for all n = 1, 2, and 3 then the sequence error for the combination shall be computed as follows:

$$eseq = ST + \sum_{n=1}^{4} SERR$$

TABLE X

EXPECTED TIME INTERVALS AND RELATIVE ENERGIES

m	Δte_m	ETH_m
1	40	42
2	43	50
3	46	19
4	44	45
5	43	33
6	48	20
7	48	41
8	46	29
9	42	29
10	43	32
11	44	23
12	52	40
13	—	20
14	—	22
15	—	41
16	—	17

Where

$$ST = \left\{ \sum_{n=1}^{3} tp(j_{n+1}) - tp(j_n) - \Delta te_{m(n)} \right\} 2/16 \quad \text{(rounded)}$$

$$SERR_n = X_n^2 / 128^2 \quad \text{(d.p.)}$$

$$X_n = \left\{ 16 * \left[128 * ep(j_n) - ETH_{m(n)} * ETOT \right] \right\} / ETOT \quad \text{(d.p., rounded)}$$

$$\text{and } ETOT = \sum_{n=1}^{4} ep(j_n)$$

If eseq can be computed and if it is less than the current value of BESTSQ (initialized as per 10.2.2.1.c), then BESTSQ shall be set to eseq and the current sequence shall be saved as the optimum thus far i.e., $OPTT_n = tp(j_n)$ and $OPTSE_n = ep(j_n)$ for n = 1, 2, 3, 4. If after all such possible combinations have been so considered and if BESTSQ resulted in a value greater than 625, all data resulting from the processing of this phrase shall be discarded, the filter gain adjustment test as per 10.2.2.3.2 shall be implemented, where this phrase shall be deemed not mis-registered, this same phrase shall be reprompted and the processing shall be repeated starting as described in 10.2.2.1. There shall be no limit to the number of times this phrase shall be reprompted in order to obtain an acceptable optimum sequence. If an acceptable optimum sequence is found, i.e., if BESTSQ $\leq$ 625, the procedure of 10.2.2.5 shall be implemented.

10.2.2.5 Creating initial reference file data. Once an acceptable optimum sequence has been obtained as per 10.2.2.4, the initial reference file data shall be generated as follows. Using the optimum sequence data $OPTT_n$ and $OPTSE_n$ for n = 1, 2, 3, 4, then, for the four words used in the phrase just prompted:

a. The pattern data for word m(n) shall be saved as follows:

$$sr_{m(n)}(i,k) = \eta(i, OPTT_n - 7 + 2k)$$

for i = 1, 2,...14
k = 1, 2,...6
n = 1, 2, 3, 4

and where $\eta(i,t)$ is the quantized data generated at time t for filter i as per 10.2.2.2.1.3.

b. The expected time intervals shall be saved as

$$s\Delta Tref_{m(n)} = OPTT_{n+1} - OPTT_n$$

for n = 1, 2, 3

c. The expected scanning error shall be set to zero: $sESE_{m(n)} = 0$

d. The number of times the reference data has been calculated shall be set to one: $NTC_{m(n)} = 1$

With this complete, the test of whether generation of initial reference data is complete shall be performed as per 10.2.2.6.

10.2.2.6 Initial reference data complete. The generation of initial reference data is not said to be complete until data has been computed for each of the possible 16 words at least one time. Therefore, if $NTC_m < 1$ for any m = 1, 2,...16 the generation of initial reference file data, 10.2.2, shall continue. If, however, $NTC_m \geq 1$ for all m = 1,2,...16, a complete set of initial reference data is available and the procedures of 10.2.3 shall be implemented. In either case, another phrase will be prompted and as such, the gain adjustment test as per 10.2.2.3.2 shall be implemented, where this phrase shall be deemed not mis-registered. Upon completion of the test, the appropriate procedure, 10.2.2 or 10.2.3, shall be implemented where the phrase to be prompted shall be the next phrase in the list created as per 10.2.1.d. If all phrases in this list have been used, a new list shall be created as per 10.1.3.3 where the first phrase in this new list shall be that which is prompted.

10.2.3 Refinement of reference file data. The refinement of reference file data consists of the processing of phrases containing the 16 words such that each word is processed at least four times. The data for each processed word is then compared with the already existing reference file data and if comparable, is used to create an average which becomes the reference file data used in subsequent verification attempts. If the processed phrase data is not comparable to the already existing reference file data, a re-intialization of the reference file data may instead become necessary and the refinement process continued until the words involved have had their reference file data updated the required four times. To perform the refinement process, it is necessary that reference file data already exist for all of the possible 16 words. To this end, the test of 10.2.2.6 shall be passed prior to starting the refinement process. The refinement process shall be as follows.

10.2.3.1 Phrase prompt initialization. Prior to the processing of speech data for the phrase selected for prompting by the previous processing, 10.2.2.6 or 10.2.3.5, the following initialization shall be implemented.

a. The parameters MIN_n, and $LTIME_n$, and $MODE_n$ used to locate valley points in the scanning error functions for the four words

quantized data $\eta(i,L)$ and the maximum filter energy e(L) for i = 1, 2,...14 corresponding to the 14 filters and corresponding to L, the time in centiseconds since the end of prompting.

10.2.3.2.2 Scanning error. For each centisecond following the end of prompting, the scanning error centered at time L - 7 for each of the four words prompted (n = 1, 2, 3, 4) shall be defined as follows:

$$SE_{\eta}(L-7) = \sum_{k=1}^{6} \sum_{i=1}^{14} X_{ik}^{2}$$

where $X_{ik} = r_{m(n)}(i,k)$ for $L+2k-14 \leq 0$

$= r_{m(n)}(i,k) - \eta(i, L+2k-14)$ for $L+2K-14>0$

$r_{m(n)}(i,k)$ is the reference pattern data, and $\eta(i,t)$ is the quantized filter data at time t.

10.2.3.2.3 Optimum sequence of scanning error valley points. To select an optimum sequence of scanning error valley points, the valley points, i.e., local minima, of each of the four scanning errors functions must be found and then, selecting one point per function, a sequence is formed. That sequence will be considered further if the time intervals between valley points compares favorably with the expected time intervals $\Delta Tref_m$ as contained in the reference data. For those sequences which do compare well, a sequence error is computed and that sequence yielding the minimum error when end-of-speech is declared is called the optimum sequence. The procedure in locating this optimum sequence shall be as follows, see Figure 8.

10.2.3.2.3.1 Locate valley points. For the current time L, each of the four scanning errors $SE_n(L - 7)$ centered at time L - 7 shall be considered in turn. For this, the nth, scanning error function, a valley locating mode $MODE_n$ and values for MIN_n and $LTIME_n$ will have been determined at time L - 1, or during phrase prompt initialization if L = 1. Based on $MODE_n$ for this function and the value of $SE_n(L - 7)$, the values to which MIN_n, $LTIME_n$ and $MODE_n$ shall be updated to be used for this function at the next centisecond, at L + 1, shall be as presented in Table XI. As an example, if $SE_n(L - 7)$ is greater than or equal to the current value of MIN_n and the current mode is search-for-peak, MIN_n shall be set to $SE_n(L - 7)$, $LTIME_n$ shall be set to L - 7 and the mode for the next time, at L + 1, shall be search-for-peak.

If the current mode is search-for-valley and $SE_n(L - 7)$ is greater than or equal to (MIN *15)/10, then a valley point is said to be

shall be initialized respectively to 900, zero and search-for-peak for all n = 1, 2, 3, 4.

b. The parameters EPEAK and SPEECH used to find the end-of-speech shall be set to zero and not-started respectively.

c. The parameter BESTSQ used to locate an optimum sequence of valley points shall be initialized to the largest positive integer possible for the computer.

d. For the words in the phrase selected for prompting at this time, the average reference data shall be computed as follows where $NTC_{m(n)}$ is the number of times data has been collected for word m to be prompted as the nth word in the phrase. The data used in the averaging, $sr_{m(n)}$ and $s\Delta Tref_{m(n)}$, are as defined in 10.2.2.5 if $NTC_{m(n)} = 1$, or are as defined in 10.2.3.4 if $NTC_{m(n)} > 1$.

(1) The reference pattern data for word m shall be averaged by:

$$r_{m(n)}(i,k) = sr_{m(n)}(i,k)/NTC_{m(n)} \quad \text{(rounded)}$$

for i = 1, 2,...14, and k = 1, 2,...6

(2) The expected time interval between word m and any next word shall be averaged by:

$$\Delta Tref_{m(n)} = s\Delta Tref_{m(n)}/NTC_{m(n)} \quad \text{(rounded)}$$

e. If the phrase to be prompted at this time is not a reprompt due to the decisions of previous processing, as is the case when refinement of reference file data first begins, the parameter NOTREG used to count misregistered phrases shall be set to zero. In the case where the current phrase is a reprompt, NOTREG shall remain unchanged.

f. The phrase selected by the previous processing shall be prompted through the prompting unit.

The processing of 10.2.3.2 shall be implemented, when this initialization is complete.

10.2.3.2 Speech processing. For each centisecond following the end of prompting, while the enrollee is repeating the prompted phrase and until end of speech is declared, 10.2.3.2.4, the procedures of this section shall be implemented.

10.2.3.2.1 Data Compression. The process of data compression as applied to the filtered data shall be the same as that which is described in 10.2.2.2.1. The output of this process is the

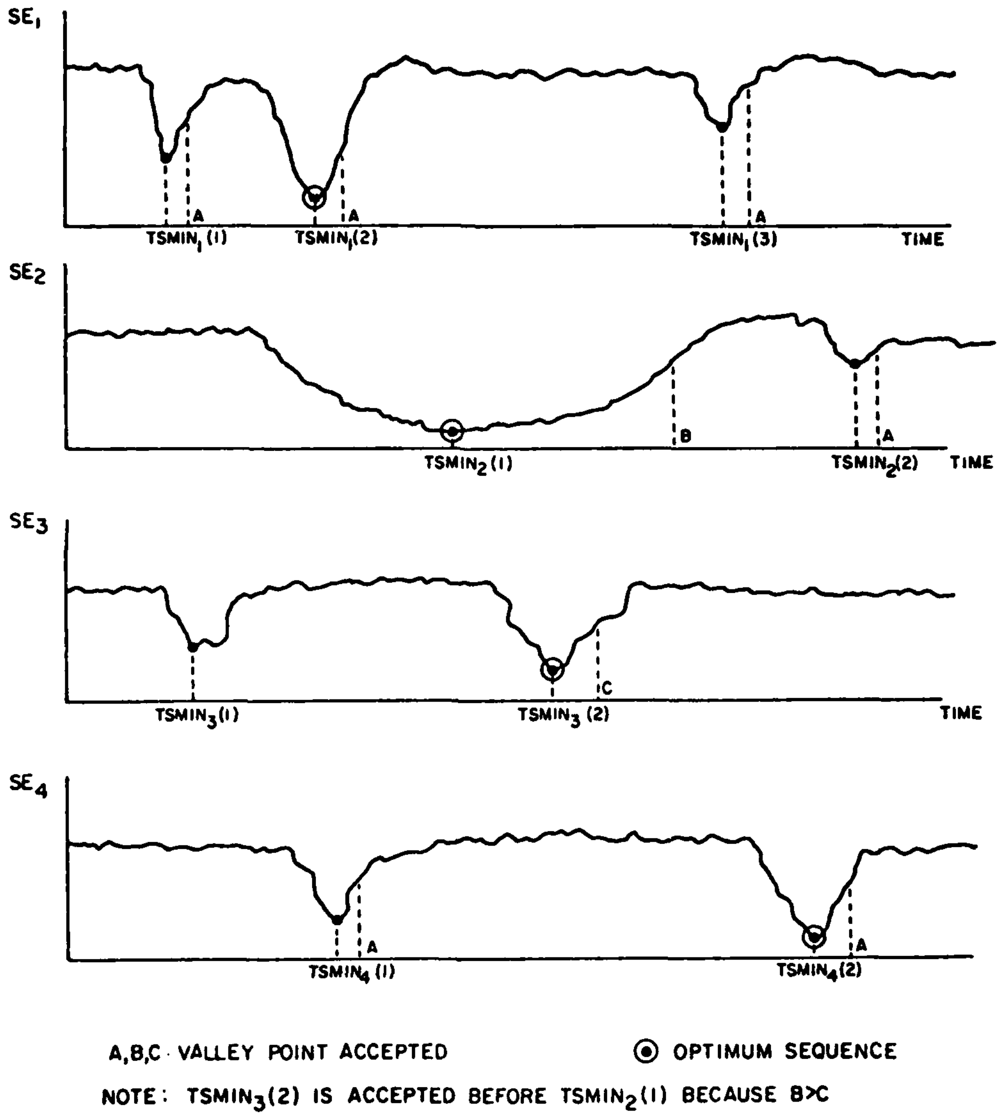


Figure 8. OPTIMUM SEQUENCE OF VALLEY POINTS

TABLE XI

VALLEY POINT TABLE

Current Mode	Condition	MIN	LTIME	MODE
Search for Peak	$SE_n(L-7) \geqslant MIN_n$	$SE_n(L-7)$	L-7	*
	$MIN_n > SE_n(L-7) > (MIN_n*10)/15$	*	*	*
	$(MIN\ *10)/15 \geqslant SE_n(L-7)$	$SE_n(L-7)$	L-7	Valley
Search for Valley	$MIN_n > SE_n(L-7)$	$SE_n(L-7)$	L-7	*
	$(MIN_n*15)/10 > SE_n(L-7) \geqslant MIN_n$	*	*	*
	$SE_n(L-7) \geqslant (MIN_n*15)/10$	$SE_n(L-7)$	L-7	Peak

*Unchanged

located. In this case, prior to the updating indicated in the table and then only if the unupdated value of MIN_n is less than or equal to 600 and the unupdated value of $LTIME_n$ is greater than seven centiseconds, the valley point shall be saved. If the valley point is to be saved, the unupdated values of MIN_n and $LTIME_n$ shall be stored in arrays $ESMIN_n$ and $TSMIN_n$ respectively where up to five such valley points shall be provided for each of the four scanning error functions. If more than five valley points are found for a given function, these arrays shall be used circularly whereby the oldest valley point is discarded in favor of the newest. If a valley point was saved at this time for this, the nth, scanning error function, the procedures of 10.2.3.2.3.2 shall be implemented next. If a valley point was not saved at this time and if $n \leq 3$, the valley locating procedures of this section shall be implemented next for the (n+1)th scanning error function. If, however, n=4 and no valley point was saved at this time, the end of speech test as per 10.2.3.2.4, or 10.3.2.3 if the verification algorithm is being implemented, shall be implemented next.

10.2.3.2.3.2 Finding the optimum sequence. If the valley point was saved as per 10.2.3.2.3.1 at the current time for the scanning error function currently under consideration (i.e., SE_n) it shall be subjected to tests to determine if it is to become a part of a sequence. If it is and if the sequence can be formed at this time, the sequence is tested to determine if it is the optimum sequence found thus far. By always retaining only the sequence which is the optimum that has been found thus far the sequence which exists when end-of-speech is declared is that sequence which becomes known as the optimum for the phrase.

In the following, the term couplet is used to denote a pair of valley points, each consisting of a time and scanning error, from consecutive scanning error functions, e.g., SE_I and SE_{I+1}, since the valley points have been saved in circular arrays $TMSIN_I$ and $ESMIN_I$, j_I is used as an index to select one of the valley points of SE_I saved in these arrays. Hence, $1 \leq I \leq 4$ but $1 \leq j_I \leq 5$. The testing procedures are as indicated below.

10.2.3.2.3.2.1 Successor couplets. Successor couplets shall be formed using the just saved valley point of SE_n and valley points already saved for SE_{n+1}. If no valley points have been saved for SE_{n+1} or if n=4 whereby SE_5 does not exist, processing shall proceed immediately to 10.2.3.2.3.2.2. If $n < 4$ and valley points have been saved for SE_{n+1}, the couplets shall be formed starting with the most recently saved valley point and progressing to the oldest saved valley point of SE_{n+1}. As each such couplet is formed it shall be tested according to 10.2.3.2.3.2.1.1 and 10.2.3.2.3.2.1.2 where I=n in these tests. When all such couplets have been so tested, processing shall proceed to 10.2.3.2.3.2.2.

10.2.3.2.3.2.1.1 Time restriction test Using the couplet selected by previous processing, the following test shall be performed. If

$$\Delta Tref_{m(I)}/2 \leq \Delta T \leq 2 * \Delta Tref_{m(I)}$$

where $\Delta Tref_{m(I)}$ is the reference data expected time interval between word m, prompted as the Ith word in the phrase, and whatever word is prompted next,

$$\Delta T = TSMIN_{I+1}(j_{I+1}) - TSMIN_I(j_I)$$

j_{I+1}, j_I are the indices to the valley points selected for this couplet,

and $TSMIN_{I+1}(j_{I+1})$ and $TSMIN_I(j_I)$ are the times of these valley points,

then the couplet is said to have passed the time restriction test. If the couplet currently under consideration passes this test, processing shall proceed to 10.2.3.2.3.2.1.1. If the couplet did not pass this test, processing shall proceed as in 10.2.3.2.3.2.1, or 10.2.3.2.3.2.2 if predecessor couplets are being tested.

10.2.3.2.3.2.1.2 Point pair test. Using the couplet which passed the time restriction test, a point pair error EW shall be computed as follows:

$$EW = \left\{\left\{\left[(MXD + 4 * \left|\Delta T - \Delta Tref_{m(I)}\right|) * (ESMIN_I\ (j_I) + 1)\right] /MXD\right\} * (ESMIN_{I+1}\ (j_{I+1}) + 1)\right\} /2048 \qquad \text{(d.p.)}$$

where ΔT, $\Delta Tref_{m(I)}$, j_I and j_{I+1} are as defined above,
$ESMIN_I\ (j_I)$ and $ESMIN_{I+1}\ (j_{I+1})$ are the scanning errors of the valley points selected for this couplet, and
MXD is the maximum of 80 and $4 * \Delta Tref_{m(I)}$.

If this couplet produces an EW greater than or equal to 70, the couplet is said to have failed the point pair test and processing shall continue as indicated in 10.2.3.2.3.2.1, or 10.2.3.2.3.2.2 for predecessor couplets. If this couplet produces an EW less than 70, the couplet along with its EW shall be saved as follows:

$$PPT_I\ (k_I, 1) = TSMIN_I\ (j_I)$$

$$PPT_I\ (k_I, 2) = TSMIN_{I+1}\ (j_{I+1})$$

$$PPE_I\ (k_I, 1) = ESMIN_I\ (j_I)$$

$$PPE_I\ (k_I, 2) = ESMIN_{I+1}\ (j_{I+1})$$

$$PPEW_I\ (k_I) = EW$$

where provision shall be made for the saving of data for up to five such couplets ($1 \leq k_I \leq 5$) for I = 1, 2, 3. If more than five couplets are found which pass these tests for any I, the associated arrays shall be used circularly whereby the oldest couplet data is discarded in favor of the newest. Having saved the couplet data, processing shall continue as in 10.2.3.2.3.2.1, or for prodecessor couplets, 10.2.3.2.3.2.2.

10.2.3.2.3.2.2 Predecessor couplets. Prodecessor couplets shall be formed, tested and saved as appropriate in the same manner as it was done for successor couplets. Predecessor couplets consist of valley points already saved for SE_{n-1} and the just saved valley point of SE_n. If no valley points have been saved for SE_{n-1} or if n=1 whereby SE_0 does not exist, processing shall proceed in accordance to 10.2.3.2.3.2.3. If n > 1 and valley points have been saved for SE_{n-1}these couplets shall be formed starting with the most recently saved valley point and progressing to the oldest saved valley point of SE_{n-1}. As each such couplet is formed, it shall be tested as in to 10.2.3.2.3.2.1.1 and 10.2.3.2.3.2.1.2 where I=N-1 in these tests. When all such couplets have been so tested, processing shall proceed to 10.2.3.2.3.2.3.

10.2.3.2.3.2.3 Forming sequence. A chain of couplets is created by finding k_1, k_2 and k_3 such that

$$PPT_I\ (k_I,\ 2) = PPT_{I+1}\ (k_{I+1},\ 1)$$

$$PPE_I\ (k_I,\ 2) = PPE_{I+1}\ (k_{I+1},\ 1)$$

for I = 1 and I = 2

where the PPT and PPE arrays are as defined in 10.2.3.2.3.2.1.2.

For each such created chain which includes both predecessor (if N > 1) and successor (if N < 4) couplets for the just saved valley point of SE_n, a sequence error, eseq, shall be computed as follows:

$$eseq = \sum_{I=1}^{3} PPEW_I(k_I)$$

Where the k_I's are the indices of the chain and PPEW is as defined in 10.2.3.2.3.2.1.2.

If this value of eseq is less than the sequence errors for all previous chains (i.e., if eseq < BESTSQ), then BESTSQ shall be set

to eseq and the valley points of the chain shall be saved as follows:

$$OPTSE_I = PPE_I(k_I, 1)$$

$$OPTT_I = PPT_I(k_I, 1)$$

$$\text{for } I = 1, 2, \text{ and } 3$$

$$\text{and } OPTSE_4 = PPE_3(k_3, 2)$$

$$OPTT_4 = PPT_3(k_3, 2)$$

These values of $OPTT_I$ (for I=1, 2, 3 and 4) shall be further adjusted based on the values of the smoothed energy function es(t) around these times. Letting successively $t_I = OPTT_I - 1$, $t_I = OPTT_I$ and $t_I = OPTT_I + 1$, the smoothed energies shall be computed as follows:

$$es(t_I) = \sum_{k=-6}^{+6} \frac{e(t_I + k)}{2}$$

where e(t) is the maximum filter energy at time t of 10.2.2.2.1.1.d.

That value of t_I which yields the maximum value of $es(t_I)$ shall be used as the reported value of $OPTT_I$. The values of $OPTSE_I$ shall not be effected by this adjustment.

When each such chain involving the just saved valley point has been so tested, or if no such chain could be created, processing shall proceed with the valley locating procedures of 10.2.3.2.3.1 if $n < 4$. If n=4, i.e., the just saved valley point was from SE_4, processing shall instead proceed to the end-of-speech test given below, or 10.3.2.3 if the verification algorithm is being implemented.

10.2.3.2.4 End-of-speech. The test for determining when end-of-speech has occurred during the refinement of reference data shall be implemented in the same manner as that described in 10.2.2.2.3. If end-of-speech is declared, no more processing of filter data is required and the data collected shall be tested for suitability as per 10.2.3.3. If end-of-speech is not declared, speech processing shall continue as per 10.2.3.2.

10.2.3.3 Suitable data. Once end-of-speech is declared, the data resulting from this processing shall be deemed to be suitable or not in the following manner. If no filter overload status condition was sensed and if an optimum sequence was found, i.e., BESTSQ was changed from its initial value of the largest positive integer, then

the data shall be deemed suitable and the reference update procedure of 10.2.3.4 shall be implemented. If, however, a filter overload occurred, the data resulting from the speech processing shall be discarded, the filter overload procedures shall be implemented and this phrase shall be reprompted as the next phrase with processing restarting as per 10.2.3.1. There shall be no limit to the number of times a phrase is reprompted in this manner due to filter overloads. If the filters did not overload but instead the phrase "mis-registered", i.e., an optimum sequence was not found, the data is unsuitable but subsequent processing is dependent on the number of times this phrase has mis-registered. In this case the parameter NOTREG shall be incremented by one. If NOTREG becomes equal to one, i.e. this is the first time the phrase mis-registered, the data resulting from the speech processing shall be discarded, the filter gain adjustment procedures of 10.2.2.3.2 shall be implemented and this phrase shall be reprompted as the next phrase with processing restarting. If, however, NOTREG becomes two, i.e. this is the second time this same phrase has mis-registered, the data compression results ($e(L)$ and $\eta(i,L)$) from 10.2.3.2.1 shall be used to re-initialize the reference data, 10.2.2.1 through 10.2.2.6 but excluding 10.2.2.1.d, no phrase shall be prompted, and all sections of 10.2.2.2.1, the compressed data is already available.

10.2.3.4 <u>Reference data update</u>. For that data deemed suitable as per 10.2.3.3 the following reference data updating shall be performed. The reference data to be updated are those values of $sr_m(i,k)$, $s\Delta Tref_m$, $sESE_m$ and NTC_m as created by 10.2.2.5 and possibly previously updated by this section. For only those words used in the prompting of this phrase, the corresponding NTC_m will determine what updating is to be performed. If for word m, prompted as the nth word in the phrase, the value of NTC_m equals 32, no updating shall be performed. If, however, NTC_m is less than 32, the following updating shall be performed using the optimum sequence data, $OPTSE_n$ and $OPTT_n$, n = 1, 2, 3, 4, defined in 10.2.3.2.3.2.3.

a. Reference pattern data

$$sr_{m(n)}(i,k) = sr_{m(n)}(i,k) + \eta(i, OPTT_n - 7 + 2k)$$

for i = 1, 2,...14, k = 1, 2,...6 and where $\eta(i,t)$ is the data generated by 10.2.3.2.1.

b. Expected time interval reference data

$$s\Delta Tref_{m(n)} = s\Delta Tref_{m(n)} + OPTT_{n+1} - OPTT_n$$

c. Expected scanning error reference data

$$sESE_{m(n)} = sESE_{m(n)} + OPTSE_n \text{ if } NTC_{m(n)} \leq 4$$

If $NTC_{m(n)} > 4$, no updating of the expected scanning error data shall be performed.

d. Data collection count

$$NTC_{m(n)} = NTC_{m(n)} + 1$$

When all indicated updating is complete, the procedures of 10.2.3.5 shall be implemented.

10.2.3.5 Refinement or termination. When updating for this phrase is complete, a determination is made as to whether further reference data refinement is required or whether the enrollment process shall be terminated. If each of the possible 16 words has a data collection count NTC_m which is greater than or equal to five, the enrollment process shall be terminated and the procedures of 10.2.4 implemented. If, however, one or more of the NTC_m is less than five, further refinement shall be required. As this will entail the prompting of another phrase, the filter gain adjustment procedures of 10.2.2.3.2 shall be implemented where the current phrase is said to be not mis-registered. Once this is complete, the processing starting at 10.2.3.1 shall be repeated where the phrase to be prompted shall be the next phrase in the list. If all 32 phrases have already been prompted, a new list shall be created in accordance to 10.1.3.3 and the first phrase in this new list shall be that which is prompted.

10.2.4 Enrollment termination. When the enrollment process is deemed complete as per 10.2.3.5, the data to be stored on the reference file associated with the individual through the coded badge, personnel code and authorization file shall be computed. The data used in this process shall be that which resulted from the last updating performed for each of the 16 words. The procedure shall be as follows:

a. Reference pattern data

$$r_m(i,k) = sr_m(i,k)/NTC_m \quad \text{(rounded)}$$

for $i = 1, 2, \ldots 14$

$k = 1, 2, \ldots 6$

$m = 1, 2, \ldots 16$

These averages shall be stored on the individual reference file providing enough storage such that updating during subsequent verifications shall permit significance to the nearest 1/32, i.e. five fractional bits shall be provided for each pattern data point

although the data at this point shall be rounded to the nearest integer.

b. Time interval reference data

$$\Delta Tref = s\Delta Tref_m/NTC_m$$

for m = 1, 2,...12

These averages shall be computed and rounded to the nearest 0.1. The results of these averages shall be stored on the individual reference file.

c. Expected scanning error reference data

$$ESE_m = sESE_m/4$$

for m = 1, 2,...16

These averages shall be computed and rounded to the nearest 0.1. Since these averages tend to be underestimated, the resultant ESE_m shall be further scaled by

$$ESE_m = (125*ESE_m)/100 \qquad \text{(d.p.)}$$

for m = 1, 2,...16

and limited to a value no less than 110. These scaled and limited averages shall also be computed and rounded to the nearest 0.1 before being stored on the individual reference file.

When all reference data has been thus computed and stored, the enrollment process is complete and the instruction "thank you" shall be prompted to the individual or enrollee.

10.2.5 Enrollment interruptions. The enrollment process shall be interruptable in either of the two following ways. If at any time during speech processing until end-of-speech is declared, the individual sends a "reprompt" signal to the processor, the current speech processing shall be halted, all intermediate results computed for the phrase shall be discarded, the same phrase shall be reprompted and the phrase processing repeated. There shall be no limit to the number of times an individual may thus request a reprompt. If at any time during the enrollment procedure the operator sends a "stop" signal to the processor, the enrollment process shall be terminated immediately and no data shall be stored. The individual shall be informed of this premature termination by an appropriate prompted instruction such as "call for assistance".

10.3 Verification algorithm. The verification algorithm consists of three phases: (1) initialization, (2) phrase processing and (3) decision making. The procedure involves the prompting of a phrase, the processing of the filtered data resulting from the entrant repeating the phrase and, if the data is suitable, the computing of a score which measures the closeness of this data to the reference file data. If the score reflects a desired degree of closeness, the entrant is said to be verified. If this is not the case, more phrase prompting may be allowed until a verified decision is made or until the entrant is said to be not verified due to exhausting the number of promptings allowed. Figure 9 is a flowchart of the verification algorithm. The notation glossary of Table VI will be adhered to in the following description.

10.3.1 Verification initialization. The verification function shall begin by the acceptance of an entrant's ID number from the pedestrian booth through the authorization file. If the entrant is authorized entry to, or exit from the restricted area, the following initialization shall be implemented.

a. The filter gain G which has possible values 1, 2, 4..., up to a maximum filter gain shall be initialized to two and the filter overload indicator shall be cleared.

b. The reference file data as initially created, 10.2.4, or as last updated, 10.3.3.4.3, shall be retrieved.

c. A list of eight phrases, two groups, of random order shall be created as per 10.1.3.3. The first group shall be called the "initial group"; the second group, the "auto-abort group". When processing begins as per 10.3.2, it shall be the first phrase in the "initial group" which shall be prompted.

d. The parameters IREG, NOTREG, EHAT and EUSE used in determining the verification decision shall be set to zero.

e. If the entrant has performed the verification process a sufficient number of times such that at least four resultant decisions have been "entrant verified", the mode of the verification algorithm shall be said to be "normal". If the entrant has not successfully verified four times, the algorithm mode shall be said to be "post-enrollment" or PE.

With this initialization complete, processing shall begin in accordance to 10.3.2.

10.3.2 Verification phrase processing. The processing of phrases during verification is identical to that of the enrollment speech processing during the refinement phase, 10.2.3.2, with the

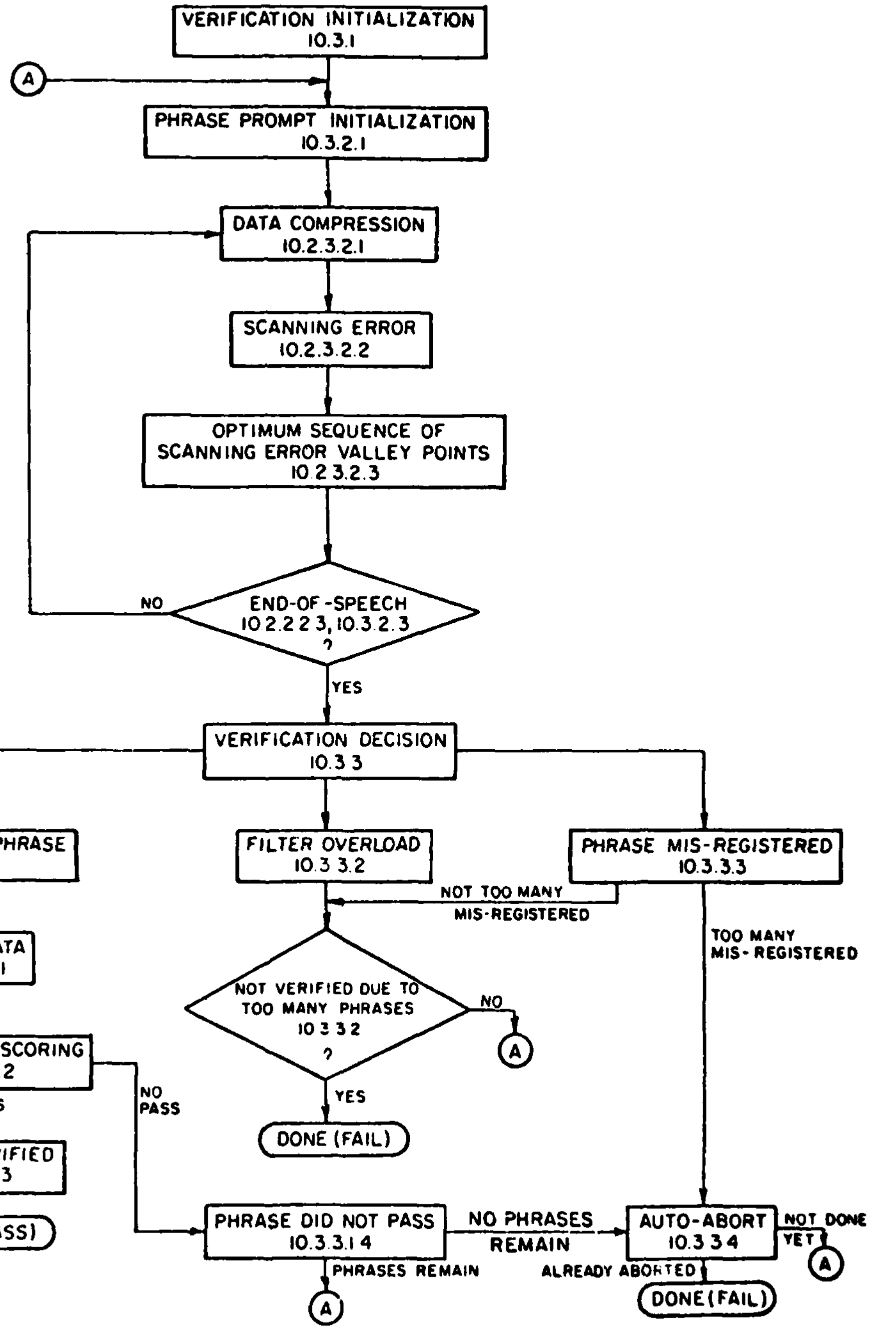


Figure 9. VERIFICATION ALGORITHM

exceptions that the phrase prompt initialization differs and the end-of-speech test involves one more step. Verification phrase processing shall be implemented as follows.

10.3.2.1 Phrase prompt initialization. The processing of speech data for the phrase selected for prompting, 10.3.1 or 10.3.3, shall be preceded by the following initialization:

a. The parameters MIN_n, $LTIME_n$, and $MODE_n$ used to locate valley points in the scanning error functions for each of the four prompted words shall be initialized respectively to 900, zero and search-for-peak for all n = 1, 2, 3, 4.

b. The parameters EPEAK, SPEECH, and MAXCUT used to find the end of speech shall be set to zero, "not-started" and 400 respectively.

c. The parameter BESTSQ used to locate the optimum sequence of valley points shall be initialized to the largest positive integer possible for the computer.

d. For the four words in the phrase selected for prompting at this time, the following summed expected scanning error shall be computed:

$$sESE = \sum_{n=1}^{4} ESE_{m(n)}$$

where $ESE_{m(n)}$ is the reference expected scanning error for word m prompted as the nth word in the phrase. This summed value shall be computed maintaining significance to the nearest 0.1 but the resultant sum shall be rounded to the nearest integer. For use only during verification phrase processing, i.e. all sections of 10.3.2, the other reference data ($r_m(i,k)$ and $\Delta Tref_m$) for the words in the phrase to be prompted shall be rounded to the nearest integer.

e. The phrase selected by the previous processing shall be prompted through the prompting unit.

When this initialization is complete, the processing of 10.3.2.2 shall be implemented.

10.3.2.2 Speech processing. For each centisecond following the end of prompting, while the entrant is repeating the prompted phrase, and until end-of-speech is declared, the speech processing procedures, 10.2.3.2, shall be implemented using the end-of-speech test which follows, 10.3.2.3, rather than that of 10.2.3.2.4.

10.3.2.3 End-of-speech. If less than 10 centiseconds has elapsed since the end of prompting, the end-of-speech test shall not be implemented and speech processing shall continue as per 10.3.2.2. If however $L \geq 10$, the test for determining when end-of-speech has occurred shall be implemented in the same manner as that described in 10.2.2.2.3 with the following addition. If end-of-speech was not declared as per 10.2.2.2.3.1 and 10.2.2.2.3.2, the following additional test shall be implemented. If an optimum sequence was found at this time as per 10.2.3.2.3.2, the value of MAXCUT shall be recomputed by:

$$MAXCUT = OPTT_1 + 7 + \sum_{n=1}^{3} \Delta Tref_{m(n)}$$

where $OPTT_1$ is the time of the first valley point in the optimum sequence and

$\Delta Tref_{m(n)}$ are the expected time intervals between the prompted words.

If the resultant value of MAXCUT is greater than 400, it shall be limited to 400. If no optimum sequence was saved at this time, MAXCUT shall be left at its previous value.

If L, the total elapsed time since the end of prompting, is greater than or equal to the current value of MAXCUT and if the maximum filter energy at this time, e(L) of 10.2.2.2.1.1, is less than EPEAK/8, with EPEAK as given in 10.2.2.2.3, then end-of-speech is declared. If end-of-speech has not been declared as per this test or those of 10.2.2.2.3, speech processing shall continue as described in 10.3.2.2. If end-of-speech is declared by any one of these tests, speech processing shall be stopped and the decison making procedures of 10.3.3 shall be implemented.

10.3.3 Verification decision. A verification decision that the entrant has successfully verified against the reference file data can only be made on those phrases which are deemed suitable.

In what follows, a phrase will be called "mis-registered" if an optimum sequence could not be found as per 10.2.3.2.3.2. If an optimum sequence could be found, the phrase will be called registered or not mis-registered.

If a filter overload condition occurred at any time during the speech processing, the data is deemed not suitable and the procedures of 10.3.3.2 shall be implemented. If no filter overload occurred but the phrase mis-registered, the data is also not suitable and the procedures of 10.3.3.3 shall be implemented. If

neither of these two conditions occurred, the data is suitable and the procedures of 10.3.3.1 shall be implemented.

10.3.3.1 Phrase registered. For those phrases which registered, the following shall be performed.

10.3.3.1.1 Saving data. For the purpose of the possible updating of reference file data and the computation of a decision score, the following shall be implemented. The count IREG of registered phrases shall be incremented by one. The words used in prompting the current phrase, the optimum sequence scanning errors, the time intervals between optimum sequence valley points, and the quantized filter data centered on these valley points shall be saved as follows:

$$uwd_{IREG}(n) = m(n) \text{ for } n = 1, 2, 3, 4$$

$$use_{IREG}(n) = OPTSE_n \text{ for } n = 1, 2, 3, 4$$

$$udt_{IREG}(n) = OPTT_{n+1} - OPTT_n \text{ for } n = 1, 2, 3, \text{ and}$$

$$uptn_{IREG}(i, k, n) = \eta(i, OPTT_n - 7 + 2k) \text{ for}$$

$i = 1, 2, \ldots 14$; $k = 1, 2, \ldots 6$; $n = 1, 2, 3, 4$;

where m(n) is the word m prompted as the nth word in the phrase, $OPTSE_n$ and $OPTT_n$ are defined in 10.2.3.2.3.2.3 and $\eta(i,t)$ is as defined in 10.2.2.2.1.3.

Additionally the parameters EHAT and EUSE shall be updated as follows:

$$EHAT = EHAT + sESE$$

$$EUSE = EUSE + \sum_{n=1}^{4} OPTSE$$

where sESE is the summed expected scanning error for the phrase and $OPTSE_n$ is as above.

Once the data has been saved in this manner, the verification scoring procedures shall be implemented.

10.3.3.1.2 Verification scoring. A decision score PDEC used to determine whether the entrant was verified on this phrase shall be computed as follows:

$$PDEC = 100*EUSE/x \qquad \text{(d.p., rounded)}$$

where x is the minimum of the product 560*IREG and the value of $\hat{e}$,

$\hat{e}$ is the maximum of the product 400*IREG and the value of EHAT, and

EUSE, IREG and EHAT are as per 10.3.3.1.1.

Based on the mode of the algorithm, normal or PE, and the group of phrases to which the current phrase belongs, initial or autoabort, the threshold corresponding to the current value of IREG as given in Table XII shall be used. If PDEC is less than or equal to this threshold, the entrant is said to have successfully verified and the procedures of 10.3.3.1.3 shall be implemented. If PDEC is greater than this threshold, the phrase did not pass and the procedures given in 10.3.3.1.4 shall be implemented.

10.3.3.1.3 Entrant verified. Only when an entrant successfully verifies as per 10.3.3.1.2 will the reference file data be updated. Using $\alpha = 2$ if the algorithm mode is normal or $\alpha = 8$ if the algorithm mode is PE, the following reference file updating shall be implemented. For each word $m = uwd_j(n)$ where $j = 1, 2, \ldots$ IREG and $n = 1, 2, 3, 4$:

$$r_m(i,k) = (32-\alpha)*r_m(i,k) + \alpha*uptn_j(i,k,n)/32$$

for i=1,2,...14 and k=1,2,...6

$$\Delta Tref_m = (32-\alpha)*\Delta Tref_m + \alpha*udt_j\ (n)/32 \qquad \text{(d.p.)}$$

$$ESE_m = (32-\alpha)*ESE_m + \alpha*use_j\ (n)/32 \qquad \text{(d.p.)}$$

where $uptn_j$, udt_j, and use_j are the data for registered phrase j, saved as per 10.3.3.1.1 and r_m, $\Delta Tref_m$ and ESE_m are the reference file data for word m with all fractional information maintained.

Once the reference file data has been updated (where r_m shall be rounded to the nearest 1/32 and $\Delta Tref_m$ and ESE_m shall be rounded to the nearest 1/10), the verification algorithm is complete and the instruction "thank you" shall be prompted to the entrant.

10.3.3.1.4 Phrase did not pass. If the phrase did not pass the verification scoring as per 10.3.3.1.2, the entrant may or may not be said to have not verified. A check shall be made to determine if more phrases remain in the list of phrases for the current group, initial or auto-abort. If there are more phrases, the verified/not verified decision shall not be made. Instead, the filter gain adjustment procedure of 10.2.2.3.2 shall be implemented and the processing shall continue as per 10.3.2 where the phrase to be prompted shall be the next phrase in the current group's list of phrases. If no more phrases remain in the current group's list, the

TABLE XII

DECISION SCORE THRESHOLDS

		Mode of Algorithm	
Phrase Group	IREG	Post-Enrollment	Normal
Initial	1	0	100
	2	0	120
	3	0	135
	4	145	145
Auto Abort	1	0	85
	2	0	110
	3	0	130
	4	145	145

auto-abort procedure of 10.3.3.4 shall instead be implemented to determine if a not verified decision is to be made.

10.3.3.2 Filter overload. If a filter overload occurred during the verificaiton phrase processing of 10.3.2, the filter overload procedures of 10.2.2.3.1 shall be implemented. Additionally, the current phrase shall be appended to the end of the list of phrases in the current group where it may or may not be reprompted as determined by subsequent decision strategy. With the appending of this phrase, the procedure of 10.3.3.2.1 shall be implemented to determine if the entrant shall be deemed not verified due to too many phrases being prompted.

10.3.3.2.1 Not verified due to too many phrases. Whenever a phrase is appended to the end of the list of phrases for the current group, a check shall be made as to whether this list has become too long. If the total number of phrases in the current group's list has reached 20 in number, including those phrases already prompted from the group, the verification process shall be halted and the decision shall be that the entrant is "not verified due to too many phrases". The reference file data shall not be updated in this case. If, however, the list is less than 20 in length, the filter gain adjustment procedure shall be performed as per 10.2.2.3.2 and the verification process shall continue from 10.3.2 where the phrase to be prompted shall be the next phrase in the current group's list.

10.3.3.3 Phrase mis-registered. For each group of phrases used in the verification process, the verification algorithm permits a maximum of one mis-registered phrase in normal mode or two mis-registered phrases in PE mode. With the current phrase being deemed mis-registered as per 10.3.3, NOTREG shall be incremented by one and if NOTREG then exceeds this maximum for the group of phrases and for the mode selected for the algorithm in 10.3.1, the auto-abort procedures of 10.3.3.4 shall be implemented. If NOTREG remains less than or equal to this maximum, then this phrase shall be appended to the end of the list of phrases for the current group and the "not verified due to too many phrases" procedure as per 10.3.3.2.1 shall be implemented.

10.3.3.4 Auto-abort. Whenever all four phrases of a group have gone through the phrase registered process, 10.3.3.1, without an entrant verified decision being made or whenever the maximum number of mis-registered phrases is exceeded as per 10.3.3.3, the procedures given herein shall be implemented. If the current group of phrases is the first or initial group, auto-abort will not have yet been implemented but shall be at this time. If, however, the current group of phrases is the second or auto-abort group, auto-abort will have already been implemented once and, since the entrant shall be permitted only two groups of phrases per verification

attempt, the entrant shall be said to be not verified. If the not verified decision is made at this time, the verification procedure shall be halted and the reference file data shall not be updated. If the decision is that auto-abort is to be implemented at this time, the following shall be performed.

a. Any updating data saved as per 10.3.3.1.1 shall be discarded.

b. The parameters IREG, NOTREG, EHAT, and EUSE shall be reset to zero.

c. The filter gain adjustment procedure of 10.2.2.3.2 shall be implemented.

d. The second or auto-abort group of phrases shall be selected for usage in subsequent processing.

When this is complete, the processing shall resume, 10.3.2, where the first phrase in this second group of phrases shall be that phrase which is prompted.

10.4 Training mode. The training mode is a special mode of enrollment used to acquaint the entrant with the enrollment process before the actual enrollment occurs. The training mode shall consist of that processing performed to create initial reference data during enrollment with the exceptions that only two groups of phrases shall be used in the processing and that the actual creation of initial reference data shall not be implemented. With these exceptions, the procedures of 10.2.1 through 10.2.2.4 shall be implemented. Throughout this processing, the enrollment interruptions given in 10.2.5 shall be permitted.

10.5 Validation Mode. The validation mode is a special mode of verification used to check out the state of the reference file data after enrollment, to assist an entrant who is exhibiting unusual difficulty. 10.3 shall apply to this mode except that the updating of 10.2.3.4 shall not be implemented. The count of the number of times the entrant has successfully verified, used to determine the verification algorithm mode, shall not be affected by this procedure.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

LMED

8